UNIVERSIDADE DE SÃO PAULO
CENTRO DE ENERGIA NUCLEAR NA AGRICULTURA

STELLA DE LIMA CAMARGO

Revealing the biosynthesis of guanitoxin, a naturally occurring neurotoxic organophosphate

Piracicaba
2020

STELLA DE LIMA CAMARGO

Revealing the biosynthesis of guanitoxin, a naturally occurring neurotoxic organophosphate

Revised version according to the Resolution CoPGr 6018 from 2011

Tese apresentada ao Centro de energia Nuclear na Agricultura da Universidade de São Paulo para obtenção do título de Doutor em Ciências

Área de Concentração: Biologia na Agricultura e no Ambiente

Orientadora: Profª. Drª. Marli de Fátima Fiore

Piracicaba

2020

Dados Internacionais de Catalogação na Publicação (CIP)

Técnica de Biblioteca - CENA/USP

De Lima Camargo, Stella

Revelando a biossíntese da guanitoxina, um organofosfato neurotóxico de ocorrência natural / Revealing the biosynthesis of guanitoxin, a naturally occurring neurotoxic organophosphate / Stella de Lima Camargo; Marli de Fátima Fiore. - - Revised version according to the Resolution CoPGr 6018 from 2011. - - Piracicaba, 2020.

109 p. : il.

Tese (Doutorado – Programa de Pós-Graduação em Ciências. Área de Concentração: Biologia na Agricultura e no Ambiente) – Centro de Energia Nuclear na Agricultura da Universidade de São Paulo, Piracicaba, 2020.

1. Cianotoxina 2. Cianobactéria 3. Atividade enzimática 4. Genoma 5. Saúde pública I.Título

CDU 577.2 (579.63 + 577.15)

Elaborada por:

Marilia R.G. Henyei
CRB-8-3631
Resolução CFB Nº 184/2017

To my husband and my parents, Fernando Camargo, Claudia and Nivaldo

ACKNOWLEDMENT

God | my lovely **husband** Nando for accepting my adventures | my **parents** Claudia for unconditional and essential love, and Nivaldo for the support always, and my **brother** Nicolas | my **family**, specially my goddaughter Analu (my little popcorn) | my **godparents** Rosangela and Marcos (*in memoriam*) | my **advisor** Dra. Marli Fiore | my **supervisor abroad** Dr. Bradley Moore | to **Cena-USP** and **Scripps-UCSD** | the **funding agencies FAPESP** (proc. 2017/06869-0) and **CAPES** | my project **collaborators** Augusto, Felipe and Ernani | my **lab friends** Ana Paula, Bruno, Danillo, Lara, Renata, Simone, Rafael Popin, Rafael Dextro, Juliana, Thierry, Carol, Endrews, Glads and Taiane | my **Moore lab friends** Taylor (my surfing buddy), Kayla (Ginger's best friend), Jon (awesome mentor I), Shaun (awesome mentor II), Anya (awesome mentor III), Kate, Trevor, Monica, Patrick, Jia Jia, Tristan (Ginger's best friend II), Tim, Percy, Dave, Charlie, Ernesto and Peter | to my **Brazilian friends in San Diego** Renata, Vinicius, Iana, Dani, Leticia, Tiago and Gabo | people from **CPG-Cena** Cleide, Gilson, Fabio, Neuda and Magali | to Dr. Márcio Dias and Fernanda for the period in his lab/São Paulo | to Dr. Daniela Trivella for the period in her lab, LNBio/CNPEM | director and all employees from Cena, speacially Marilia from library | my **lovely friends** Nathalia, Nayara and Roberta | to all people that somehow helped me to grow professionally and personally during this path. Thank you.

Stella de Lima Camargo

"A wise man is one who admits that he does not know something, because if a man does not know what a thing is, it is already an advance of knowledge. However, if he does not know, but pretends to know, he is mistaken and goes backwards, inhibiting the search for knowledge"

*(**Carl Jung** | 1875 – 1961)*

ABSTRACT

DE LIMA CAMARGO, STELLA. **Revealing the biosynthesis of guanitoxin, a naturally occurring neurotoxic organophosphate**. 2020. 109 p. Tese (Doutorado em Ciências) – Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, 2020.

Sphaerospermopsis torques-reginae is a planktonic freshwater cyanobacteria species with broad distribution in several Brazilian aquatic environments. Some strains of this species are known to produce the neurotoxic organophosphate guanitoxin (GNT) as well as spumigin and anabaenopeptin, both protease inhibitors. GNT is a powerful irreversible inhibitor of the acetylcholinesterase enzyme which acts in the peripheral nervous cells of mammals. While GNT has been known for decades, the genes encoding its biosynthetic pathway are unknown. The sequencing and assembly of the *S. torques-reginae* ITEP-024 5.2 Mbp genome allowed us to identify a 12.5 kbp gene cluster potentially involved in GNT production. To rigorously validate our genomic finding, we over-expressed, purified and characterized many of the GNT biosynthetic enzymes *in vitro*, confirming our proposed biosynthetic cluster. Also, a vector was designed that assembled two of the biosynthetic genes with the goal to produce the intermediate enduracididine, a rare amino acid. The combination of *in vivo* and *in vitro* experiments successfully validated our proposed gene cluster. Using our results, we showed that this biosynthetic gene cluster was actively expressed in one of the largest cyanobacterial blooms hotspots, Lake Erie, Toledo, USA. This suggest that the potent neurotoxin guanitoxin could be in drinking water supplies around the world. Worryingly, guanitoxin is often overlooked and rarely tested for in reservoirs in spite of the danger it poses. This study contributed to the characterization of the GNT biosynthetic pathway, which may generate a patent enabling the development of rapid and sensitive methods for the detection and monitoring of this toxin in freshwaters used for public supply. Moreover, this work may offer potential for the development of new neurological medicines and tools.

Keywords: Cyanotoxin. Cyanobacteria. Enzyme activity. Genome. Public health.

RESUMO

DE LIMA CAMARGO, STELLA. **Revelando a biossíntese da guanitoxina, um organofosfato neurotóxico de ocorrência natural**. 2020. 109 p. Tese (Doutorado em Ciências) – Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, 2020.

Sphaerospermopsis torques-reginae é uma espécie planctônica de cianobactéria de água doce com ampla distribuição em ambientes aquáticos brasileiros. Sabe-se que algumas linhagens desta espécie produzem o organofosforado neurotóxico guanitoxina (GNT), bem como espumigina e a anabaenopeptina, ambos inibidores da protease. GNT é um potente inibidor irreversível da enzima acetilcolinesterase, atuando em células nervosas periféricas de mamíferos. Embora GNT seja conhecido há décadas, os genes que codificam sua via biossintética são desconhecidos. O sequenciamento e montagem do genoma da cianobactéria *S. torques-reginae* ITEP-024 de tamanho 5,2 Mbp, nos permitiu identificar um agrupamento gênico de 12,5 kbp potencialmente envolvido na produção de GNT. Para validar nosso achado genômico a maioria das enzimas biossintéticas foram superexpressadas, purificadas e caracterizadas *in vitro*, confirmando o agrupamento gênico proposto. Além disso, um vetor foi desenvolvido com os dois primeiros genes biossintéticos com o objetivo de produzir o intermediário enduracididina, um raro aminoácido. A combinação de experimentos *in vivo* e *in vitro* validou com sucesso nosso agrupamento gênico. Usando nossos resultados, mostramos que esse agrupamento gênico foi expresso ativamente em um dos maiores locais do mundo de florações de cianobactérias, o Lago Erie, Toledo, EUA. Isso sugere que essa potente neurotoxina, guanitoxina, pode estar em reservatórios de água potável em todo o mundo, usados para abastecimento público. O preocupante é que a guanitoxina é frequentemente negligenciada e raramente testada em reservatórios, apesar do perigo que representa. Este estudo contribuiu para a caracterização da via biossintética da GNT, podendo gerar uma patente que permita o desenvolvimento de métodos rápidos e sensíveis para a detecção e monitoramento dessa toxina em águas doces utilizadas para abastecimento público. Além disso, este trabalho pode oferecer potencial para o desenvolvimento de novos medicamentos neurológicos.

Palavras-chaves: Cianotoxina. Cianobactéria. Atividade enzimática. Genoma. Saúde pública.

SUMMARY

1 INTRODUCTION

Cyanobacteria and harmful blooms

Cyanobacteria are an ancient and diverse group of bacteria that exist in a wide variety of habitats including marine and freshwater, as well as extreme environments such as frozen lakes, hot springs and soda lakes. As a Gram-negative oxygenic photoautotrophic microorganism, many cyanobacteria have the ability to perform both nitrogen and carbon fixation, where they play an important role in the carbon and nitrogen global cycles.

Cyanobacteria are also known to produce toxins and form toxic blooms in freshwater bodies all over the world, becoming an important human health concern. Cyanobacteria have developed different and highly effective ecophysiological adaptations and strategies to survive and dominate in both natural and environments modified by anthropogenic action. For example, *Dolichospermum, Cylindrospermopsis, Microcystis* and *Planktothrix* are dominant genera capable of persisting in high nutrients concentrations, low nitrogen-to-phosphorus ratios, low light levels, reduced mixing, and high temperatures (PAERL; HUISMAN, 2009; PAERL; PAUL, 2012).

In summer season, cyanobacterial toxic harmful blooms (CyanoHBs) occur worldwide. In particular, Lake Erie, a member of the US-Canada Great Lakes, has experienced increases in frequency, intensity and duration of CyanoHBs in the last few decades (PAERL et al., 2011). The economic cost of the blooms can be enormous in terms of water quality, recreational use, fishing, and property values. Removal of cyanobacteria and their toxins from drinking water reservoirs can also lead to high water treatment costs (BURFORD et al., 2020). When researchers and water managers plan to deal with long-term control of CyanoHBs, they usually face two challenges: 1) nutrient over-enrichment in watersheds modified by human action, and 2) climate changes, especially global warming and alteration of rainfalls, which induce a modification of droughts and floods patterns (PAERL et al., 2011). While increased nutrient loading, mostly nitrogen and phosphorus, is linked to dominance of toxic and nontoxic strains of cyanobacteria, the biochemical pathways potentially important for the induction of blooming events are still poorly understood (HARKE et al., 2016).

Increasingly, harmful blooms are being reported worldwide and the effects in the environment are notable. Those include depleting oxygen in the water, what can cause fish kills, as well as blocking sunlight from reaching organisms deeper in the water column, and environmental degradation due to high biomass formed during the blooms. Secondary effects such as alteration in the food web efficiency, and changes in nutrient chemistry in bodies of

water, can also in long term affect the diversity and abundance of other species in the aquatic environment (BURFORD et al., 2020).

The complexity of climate change and ecosystem interactions makes CyanoHBs highly challenging. Nonetheless, the magnitude of these events and their consequences in the future makes this topic highly deserving of scientific attentions. Novel approaches for long-term control of CyanoHBs need to be applied including not only control of the factors that lead to blooms, such as increasing global temperatures or anthropogenic nutrient loading, but also a detailed understanding of the cyanobacterial strains that forming harmful blooms (HO et al., 2019, URBAN et al., 2020, BURFORD et al., 2020, RALSTON et al., 2020).

Cyanobacteria natural products – guanidine compounds

The chemical diversity of cyanobacterial natural products involves more than 1100 secondary metabolites with unique chemical structures, many of which have interesting biological activities (DITTMANN et al., 2015). Those compounds can be comprised into a diverse collection of structural classes including peptides, polyketides, alkaloids, lipids and terpenes, and amongst them guanidine compounds.

The phylum cyanobacteria is one of the main sources for natural guanidine, molecules with high diversity of structure and biological activities, which are particularly appropriate compounds for drug developments due to their increase hydrophilic nature (BERLINCK et al., 2017). Natural guanidine isolated from cyanobacteria include cylindrospermopsin, variants of saxitoxin and microcystin, as well as derivatives of spumigins and guanitoxin, formerly known as anatoxin-a(S) (FIORE et al., 2020) (Figure 1). These compounds are often a public health concern due to their toxicity against mammals and other classes of animals and plants.

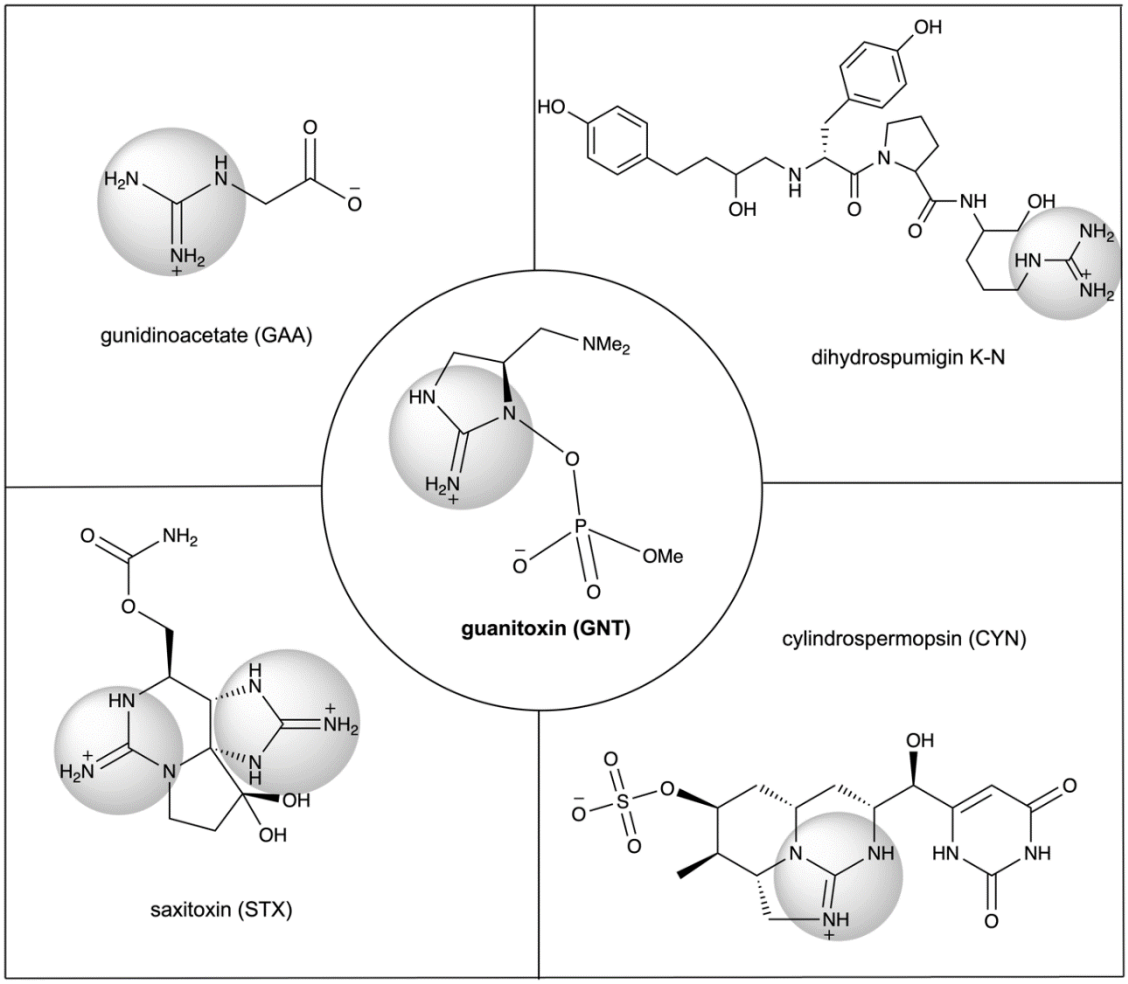


Figure 1 – Natural guanidine compounds produced by cyanobacteria. Guanidine moiety is shown in shaded gray spheres

Human health is at risk when toxic cyanobacteria blooms develop in drinking water supplies and recreational waterbodies, because those toxins are among the most powerful poisons known and these toxins have no known antidotes. The most prevalent cyanotoxins are microcystins, a family of hepatotoxic heptapeptides that have been implicated in the last few decades in fatal intoxications of humans and livestock. For example, the fatalities attributed to microcystin occurred at a hemodialysis clinic in Caruaru, Brazil, where at least 50 patients were intoxicated and showed symptoms of acute liver damage after contact with contaminated water (JOCHIMSEN et al., 1998; VAN HALDEREN et al., 1995; FRAZIER et al., 1998). Microcystins inhibit eukaryotic protein phosphatase type 1 and 2a, becoming able to access liver cells by active transport (PEARSON et al., 2016). One of the most recent case of high levels of microcystin contamination was reported in the city of Toledo, Lake Erie - OH, US, where the authorities had shut down the drinking water supply to the city (CHA; STOW, 2015).

Saxitoxin, the most studied member of the paralytic shellfish toxins (PSTs), is a potent neurotoxic alkaloid belonging to family of 57 analogs. Intoxication with PSTs can result in the severe and occasionally fatal illness know as paralytic shellfish poison (PSP) or saxitoxin pufferfish poison (SPFP). This intoxication is caused by the reversible intermolecular interaction between the positive charge of saxitoxin in the guanidinium moiety and negative charge of carboxyl groups at Na^+ channels in an equimolar rate, blocking the channels (WIESE et al., 2010). The unique to this toxin is that can be produced by both marine eukaryotic dinoflagellates and freshwater cyanobacteria. In the past, high levels of saxitoxin was detected in freshwater supplies in Australia, Brazil, USA, Mexico, Germany and China (CARMICHAEL et al., 1997; HOEGER et al., 2004; CLEMENTE et al., 2010; LIU et al., 2006; BERRY; LIND et al., 2010; BALLOT et al., 2010; CODD, 1995; SIVONEN; JONES et al., 1999). It has lead to the death of marine life and livestock, and the case of two fishermen that died after ingesting the filter feeding bi-valve *Aulacomya ater* in the Chilean Patagonian Fjords (GARCIA et al., 2004; GUY et al., 2009; STEWART et al., 2008).

One of the most diverse toxin, in terms of cyanobacteria producer, cylindrospermopsin is an extracellular polyketide-derived alkaloid known to present hepatoxicity, nephrotoxicity and general cytotoxicity effects, as well as a potent protein synthesis inhibitor (PEARSON et al., 2016). This toxin was originally isolated from an outbreak that occurred in Australia in 1979 where the majority of cases were children that showed a severe hepatoenteritis (HAWKINS et al., 1997). The toxin has since been detected in many cyanobacteria around the globe (PEARSON et al., 2016).

Finally, anatoxin-a is a neurotoxic alkaloid known to quickly intoxicate animals through its action as an antagonist of the nicotinic acetylcholine receptor (GORHAM et al., 1964). After a cattle poisoning incident in 1961, this compound was isolated for the first time in 1970s from the cyanobacteria *Anabaena flos-aquae* NRC-44 (DELVIN et al., 1977).

Harmful blooms impacts are not predictable as those from conventional chemical contaminants in waterbodies. The presence of HAB toxins can negatively influence water quality monitoring, assessment, and managements practice. It can also have a significant economic impact in the society, including as closure of fisheries and recreational areas. Last but not least, currently we understand best the effects of acute toxicity, but almost nothing is known about chronic subacute levels of toxin exposure.

Guanitoxin – natural organophosphate

Unlike the previously mentioned cyanobacterial toxins that have a known biosynthetic gene cluster and are actively monitored for in the environment, another potent cyanobacteral neurotoxin, guanitoxin (GNT), is not. Formerly named anatoxin-a(S), GNT is a natural organophosphate (FIORE et al., 2020), unlike other cyanobacterial neurotoxins (saxitoxin, anatoxin-a and homoanatoxin-a), which are alkaloids. GNT was discovered in the *Anabaena flos-aquae* NRC 525-17 strain, isolated from a bloom in Buffalo Lagoon, Saskatchewan Province, Canada in 1965 (CARMICHAEL; GORHAM, 1978). Its chemical structure has been described as a methyl phosphate ester of a cyclic *N*-hydroxyguanidine ($C_7H_{17}N_4O_4P$) of molecular weight 252 Da (MATSUNAGA et al., 1989). Pharmacological studies have shown that animals exposed to GNT showed characteristic symptoms of excessive cholinergic stimulation such as salivation, tearing, urinary incontinence, muscle fasciculation and respiratory paralysis (MAHMOOD; CARMICHAEL, 1986). Subsequently, it has been shown to be a potent and irreversible acetylcholinesterase inhibitor acting as the same way as organophosphate toxic compounds (COOK et al., 1988; MAHMOOD; CARMICHAEL, 1987; PETRIKOVICS et al., 2004; PITA et al., 2003; TUOVINEN et al., 1999; VAN APELDOORN, 2007).

The occurrence of GNT has been reported in countries across the world including Canada, Denmark, the United States, and Brazil. The *Anabaena lemmermannii* PH-160, isolated from a lake in Denmark, has been implicated in the death of birds (HENRIKSEN et al., 1997; ONODERA et al. 1997). In the United States, guanitoxin-associated toxicosis has been reported in dogs in South Dakota (MAHMOOD et al., 1998) and in pigs and ducks in Illinois (COOK et al., 1988). In Brazil, acetylcholinesterase inhibition activity was observed in extracts of *Sphaerospermopsis torques-reginae* strains ITEP-024, ITEP-025 and ITEP-026, isolated from Tapacurá reservoir blooms in Recife, PE (MOLICA et al., 2005). Subsequently, the production of GNT by these three *S. torques-reginae* strains was confirmed by mass spectrometry (DÖRR et al., 2010).

In 1992, work on zwitterionic structure of guanitoxin, led to a few deductions about the origin of carbons. The carbon scaffold of GNT were found to be derived from an amino acid, in which the carbons of the central structure of tri-amino propane and the guanidine group were derived from L-arginine (MOORE et al., 1992). In addition, two putative intermediates were observed, (2*S*, 4*S*)-4-hydroxyarginine and enduracididine (HEMSCHEIDT et al., 1995). The biosynthesis of this organophosphate, therefore, likely differs from the other cyanotoxins already described, whose biosynthesis is accomplished by the action of Non Ribosomal Peptide

Synthetases (NRPSs) or Polyketide Synthases (PKSs) modular enzymatic assembly line pathways. Saxitoxin, for example, also has an arginine as precursor. Its biosynthetic pathway uses a single module PKS to assembly a linear precursor from a coenzyme A (malonyl-CoA), an arginine and a methyl group derived from SAM (CHUN et al., 2018). In contrast, GNT was hypothesized to result from the modifications undergone by the precursor L-arginine, leading to the final structure of the toxin.

Several natural products pathways have enzymes that interact with L-arginine to produce molecules similar to those in the proposed GNT pathway (MOORE et al., 1993). The cyclic guanidinium forming enzymes of the viomycin (Vio), enduracidine (End), and mannopeptimycin (Mpp) biosynthetic pathways can work as a guide to map guanitoxin biosynthesis (HELMETAG et al., 2009; THOMAS et al., 2003; BURROUGHS et al., 2013; HAN et al., 2015; HALTLI et al., 2005, DUNHAM et al., 2018, HEDGES et al., 2018, HAN et al., 2018, HEDGES; RYAN, 2019) .

Radiolabeling studies from 1980s investigated the biosynthesis of enduracididine from enduracidin pathway, showing that arginine, ornithine and citrulline were found to be precursors (HATANO et al., 1984). Analyzing enduracidin and mannopeptimycin biosynthetic gene clusters, three pairs of genes responsible for enduracididine biosynthesis share a high nucleotide sequence homology, *mppP*/*endP*, *mppR*/*endR*, and *mppQ*/*endQ* (Figure 2). MppP is an unusual PLP dependent enzyme that uses molecular oxygen derived from water to catalyze the oxidation of L-arginine to form 2-oxo-4(S)-hydroxy-5-guanidinovaleric acid (HAN et al., 2018). MppR, an acetoacetate decarboxylase-like enzyme, carries out the next step in the pathway by cyclizing the product of MppP. Finally, an aminotransfer reaction PLP dependent proposed to be completed by MppQ forms free L-enduracididine. In mannopeptimycins biosynthesis L-enduracididine will be β-hydroxylated by the Fe(II), α-ketoglutarate dependent oxygenase, MppO (Figure 2). Besides enduracidin and mannopeptimycins biosynthesis, teixobactin antibiotic (OPPEDIJK et al., 2015) and enduracyclinones aromatic polyketide also contain enduracididine attached in a fused six ring carbon skeleton (MONCIARDINI et al., 2019).

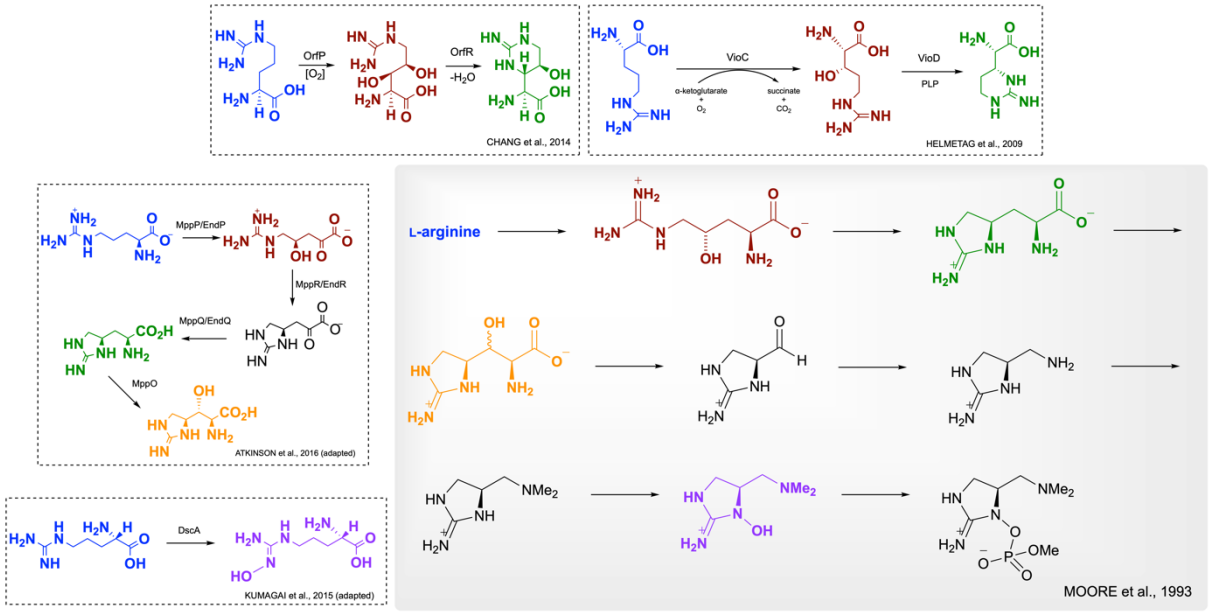


Figure 2 – Scheme showing related enzymes for mapping guanitoxin biosynthetic pathway

Heterologous expression systems as a method to explore biosynthetic pathways of cyanobacteria

Cyanobacteria are promising sources for new and exotic natural products, some of which are powerful and potent toxins. Due to the toxicity and interesting biochemical profile of those compounds, cyanobacterial biosynthetic gene cluster have been the subject of detailed biochemical study to understanding the complex biosynthesis associated with cyanotoxins.

Pioneering studies on cyanotoxins biosynthesis were typically based in radio labelled intermediate feeding experiments, and analyzing their incorporation into cyanobacterial products, or total synthesis of precursors and final product (MOORE et al., 1992, HUMPHREY et al., 1996). With the advancement of DNA sequence technologies, genomics has become a rapid approach to discover putative cyanotoxin biosynthetic gene clusters. They can now be combined with heterologous expression of genes and mutagenesis to rigorously link gene clusters to natural products.

One of the main bottlenecks of heterologous expression of biosynthetic gene clusters (BGC) is the compatibility between native organism and expression host. It is generally believed that heterology expression of BGCs is most successful when the production host is closely related to the native organism. However, even with most similar hosts, the yield of production can still be very low, pointing to more a complex machinery of natural product production that are not directly encoded by gene cluster (ZHANG et al., 2017).

Integrating enzymes heterologous expression to validate biosynthetic steps into microbial chassis (e.g. *Escherichia coli*), can be a smart tool to confirm BGC, linked with

synthesis and chemoenzymatic synthesis. The use of *in vitro* biochemical techniques to characterize biosynthetic enzymes in cyanotoxins pathways, could cover the lack of genetically manageable cyanobacteria strains and the difficult to use cyanobacteria for cloning and transformation applications (CULLEN et al., 2018).

1.1 Hypothesis

In the genome of *Sphaerospermopsis torques-reginae* ITEP-024 was identified a biosynthetic gene cluster which encodes ten biosynthetic enzymes responsible for the modification of the precursor L-arginine to produce the cyanobacterial neurotoxin, guanitoxin (GNT).

1.2 Objectives

1.2.1 General objectives

The general goal of this thesis is the confirmation of the gene cluster responsible for the production of the noxious toxin, guanitoxin.

1.2.2 Specific objectives

To achieve the general objective of this thesis the following specifics objectives were considered:

- Expression of single biosynthetic enzymes of the pathway in a well know heterologous system (*E. coli*) to evaluate proposed activity and mechanisms of each biosynthetic step by enzymatic assays (*in vitro* confirmation);
- Usage of heterologous production systems for the expression of the entire biosynthetic gene cluster in non-native host;
- Evaluate the production of the intermediates by analytical analysis.

1.3 Structural of the thesis

This thesis contains an introductory initial section followed by two studies presented in scientific manuscript format written in English language. One of the studies was published in index journal. The supplementary material indicated in each chapter are available in the Appendix section as well as a study published during this doctoral period referring to the work conducted previously in the master's degree.

References

BALLOT, A.; FASTNER, J.; WIEDNER, C. Paralytic shellfish poisoning toxin-producing cyanobacterium *Aphanizomenon gracile* in Northeast Germany. **Applied Environmental Microbiology**, Washington, DC, v. 76, p. 1173–1180, 2010.

BERLINCK, R.G.S.; BERTONHA, A.F.; TAKAKI, M.; RODRIGUEZ, J.P.G. The chemistry and biology of guanidine natural products. **Natural Products Report**, London, v. 34, p. 1264-1301, 2017.

BERRY, J.P.; LIND, O. First evidence of "paralytic shellfish toxins" and cylindrospermopsin in a Mexican freshwater system, Lago Catemaco, and apparent bioaccumulation of the toxins in "tegogolo" snails (*Pomacea patula catemacensis*). **Toxicon**, Oxford, v. 55, p. 930–938, 2010.

BURFORD, M.A.; CAREY, C.C.; HAMILTON, D.P.; HUISMAN, J.; PAERL, H.W.; WOOD, S.A.; WULFF, A. Perspective: Advancing the research agenda for improving understanding of cyanobacteria in a future of global change. **Harmful Algae**, Amsterdam, v. 91, 2020. doi: 10.1016/j.hal.2019.04.004.

BURROUGHS, A.M.; HOPPE, R.W.; GOEBEL, N.C.; SAYYED, B.H.; VOEGTLINE, T.J.; SCHWABACHER, A.W.; ZABRISKIE, T.M.; SILVAGGI, N.R. Structural and functional characterization of MppR, an enduracididine biosynthetic enzyme from *Streptomyces hygroscopicus*: functional diversity in the acetoacetate decarboxylase-like superfamily. **Biochemistry**, Washington, DC, v. 52, p. 4492-4506, 2013.

CARMICHAEL, W.W.; EVANS, W.R.; YIN, Q.Q.; BELL, P.; MOCZYDLOWSKI, E. Evidence for paralytic shellfish poisons in the freshwater cyanobacterium *Lyngbya wollei* (Farlow ex Gomont) comb.nov. **Applied Environmental Microbiology**, Washington. DC, v. 63, p. 3104–3110, 1997.

CARMICHAEL, W.W.; GORHAM, P.R. Anatoxins from clones of *Anabaena flos-aquae* isolated from lakes of Western Canada. **Mitteilungen Internationale Vereinigung für Theoretische und Angewandte Limnologie**, Heidelberg, v. 21, p. 285-295, 1978.

CHA, Y., STOW, C.A. Mining web-based data to assess public response to environmental events. **Environmental Pollution**, London, v. 198, p. 97–99, 2015.

CHUN, S.W.; HINZE, M.E.; SKIBA, M.A.; NARAYAN, A.R.H. Chemistry of a Unique Polyketide-like Synthase. **Journal of the American Chemical Society**, Washington, DC, v. 140, n. 7, p. 2430-2433, 2018.

CLEMENTE, Z.; BUSATO, R.H.; OLIVEIRA RIBEIRO, C.A.; CESTARI, M.M.; RAMSDORF, W.A.; MAGALHÃES, V.F.; WOSIACK, A.C.; SILVA DE ASSIS, H.C. Analyses of paralytic shellfish toxins and biomarkers in a southern Brazilian reservoir. **Toxicon**, Oxford, v. 55, p. 396–406, 2010.

CODD, G.A. Cyanobacterial toxins: occurrence, properties and biological significance. **Water Science and Technology**, London, v. 32, p. 149–156, 1995.

COOK, W.O.; BEASLEY, V.R.; DAHLEM, A.M.; DELLINGER, J.A.; HARLIN, K.S.; CARMICHAEL, W.W. Comparison of effects of anatoxin-a(S) and paraoxon, physostigmine and pyridostigmine on mouse brain cholinesterase activity. **Toxicon**, Oxford, v. 26, n. 8, p. 750–753, 1988.

CULLEN, A.; PEARSON, L.A.; MAZMOUZ, R.; LIU, T.; SOERIYADI, A.H.; ONGLEY, S.E.; NEILAN, B.A. Heterologous expression and biochemical characterization of cyanotoxin biosynthesis pathways. **Natural Product Reports**, London, v. 36, p. 1117-1136, 2019.

DEVLIN, J.P.; EDWARDS, O.E.; GORHAM, P.R.; HUNTER, N.R.; PIKE, R.K.; STAVRIC, B. Anatoxin-a, a toxic alkaloid from *Anabaena flos-aquae* NRC-44h. **Canadian Journal of Chemistry**, Ottawa, v. 55, p. 1367–1371, 1977.

DITTMANN, E.; GUGGER, M.; SIVONEN, K.; FEWER, D.P. Natural product biosynthetic diversity and comparative genomics of the cyanobacteria. **Trends in Microbiology**. Cambridge, v. 23, n. 10, p. 642-652, 2015.

DÖRR, F.A.; RODRÍGUEZ, V.; MOLICA, R.; HENRIKSEN, P.; KROCK, B.; PINTO, E. Methods for detection of anatoxin-a(S) by liquid chromatography coupled to electrospray ionization-tandem mass spectrometry. **Toxicon**, Oxford, v. 55, n. 1, p. 92–99, 2010.

DUNHAM, N.P.; CHANG, W.; MITCHELL, A.J.; MARTINIE, R.J.; ZHANG, B.; BERGMAN, A.J.; RAJAKOVICH, L.J.; WANG, B.; SILAKOV, A.; KREBS, C.; BOAL, A.K.; BOLLINGER, J.M. Two distinct mechanisms for c−c desaturation by iron(ii)- and 2-(oxo)glutarate-dependent oxygenases: importance of α-heteroatom assistance. **Journal of the American Chemical Society**. Washington, DC, v. 140, p. 7116−7126, 2018.

FIORE, M.F.; LIMA, S.T.; CARMICHAEL, W.W.; MCKINNIE, S.M.K.; CHEKAN, J.R.; MOORE, B.S. Guanitoxin, re-naming a cyanobacterial organophosphate toxin. **Harmful Algae**, Amsterdam, v. 92, e101737, 2020.

FRAZIER, K.; COLVIN, B.; STYER, E.; HULLINGER, G.; GARCIA, R. Microcystin toxicosis in cattle due to overgrowth of blue-green algae. **Veterinary and Human Toxicology**, Manhattan, v. 40, n. 1, p. 23-24, 1998.

GARCIA, C.; DEL CARMEN BRAVO, M.; LAGOS, M.; LAGOS, N. Paralytic shellfish poisoning: Post-mortem analysis of tissue and body fluid samples from human victims in the Patagonia fjords. **Toxicon**, Oxford, v. 43, p. 149–158, 2004.

GORHAM, P.R.; MCLACHLAN, R.W.; HAMMER, U.T. Isolation and culture of toxic strains of *Anabaena flos-aquae*. **Verhandlungen der Internationalen Vereinigung für Theoretische und Angewandte Limnologie**, Stuttgart, v. 19, p. 796–804, 1964.

GUY, A.L.; GRIFFIN, G. Adopting alternatives for the regulatory monitoring of shellfish for paralytic shellfish poisoning in Canada: interface between federal regulators, science and ethics. **Regulatory Toxicology and Pharmacology**, Amsterdam, v. 54, p. 256–263, 2009.

HALTLI, B.; TAN, Y.; MAGARVEY, N.A.; WAGENAAR, M.; YIN, X.; GREENSTEIN, M.; HUCUL, J.A.; ZABRISKIE, T.M. Investigating β-hydroxyenduracididine formation in the biosynthesis of the mannopeptimycins. **Chemistry & Biology**, New York, v. 12, p. 1163-1168, 2005.

HAN, L.; SCHWABACHER, A.W.; MORAN, G.R.; SIVAGGI, N.R. *Streptomyces wadayamensis* MppP is a pyridoxal 5′-phosphate-dependent L-arginine α-deaminase, γ-hydroxylase in the enduracididine biosynthetic pathway. **Biochemistry**, Washington, DC, v. 54, p. 7029-7040, 2015.

HAN, L.; VUKSANOVIC, N.; OEHM, S.A.; FENSKE, T.G.; SCHWABACHER, A.W.; SILVAGGI, N.R. *Streptomyces wadayamensis* mppP is a PLP-dependent oxidase, not an oxygenase. **Biochemistry**, Washington, DC, v. 57, n. 23, p. 3252-3264, 2018.

HARKE, M.J.; DAVIS, T.W.; WATSON, S.B.; GOBLER, C.J. Nutrient-Controlled Niche Differentiation of Western Lake Erie Cyanobacterial Populations Revealed via Metatranscriptomic Surveys. **Environmental Science & Technology**, Washington, DC, v. 50, n. 2, p. 604-615, 2016.

HATANO, K.; NOGAMI, I.; HIGASHIDE, E.; KISHI, T. Biosynthesis of enduracidin: origin of enduracididine and other amino acids. **Agricultural and Biological Chemistry**, Tokyo, v. 48, p. 1503–1508, 1984.

HAWKINS, P.R.; CHANDRASENA, N.R.; JONES, G.J.; HUMPAGE, A.R.; FALCONER, I.R. Isolation and toxicity of *Cylindrospermopsis raciborskii* from an ornamental lake. **Toxicon**, Oxford, v. 35, n. 3, p. 341–346, 1997.

HEDGES, J.B.; KUATSJAH, E.; DU, Y.L.; ELTIS, L.D.; RYAN, K.S. Snapshots of the catalytic cycle of an O_2, pyridoxal phosphate-dependent hydroxylase. **ACS Chemical Biology**, Washington, DC, v. 13, n. 4, p. 965-974, 2018.

HEDGES, J.B.; RYAN, K.S. Biosynthetic Pathways to Nonproteinogenic α-Amino Acids. **Chemical Reviews**, Washington, DC, 2019. doi: 10.1021/acs.chemrev.9b00408. Accessed at: Mar. 17, 2020.

HELMETAG, V.; SAMEL, S.A.; THOMAS, M.G.; MARAHIEL, M.A.; ESSEN, L. Structural basis for the erythro-stereospecificity of the L-arginine oxygenase VioC in viomycin biosynthesis. **FEBS Journal**, Oxford, v. 276, p. 3669-3682, 2009.

HEMSCHEIDT, T.; BURGOYNE, D. L.; MOORE, R.E. Biosynthesis of anatoxin-a(S). (2S,4S)-4-hydroxyarginine as an intermediate. **Journal of the Chemical Society, Chemical Communications**. London, p. 205-206, 1995.

HENRIKSEN, P.; CARMICHAEL, W.W.; NA, J.; MOESTRUP, O. Detection of an anatoxin-a(S)-like anticholinesterase in natural blooms and cultures of cyanobacteria/blue-green algae from Danish lakes and in the stomach contents of poisoned birds. **Toxicon**, Oxford, v. 35, n. 6, p. 901-913, 1997.

HO, J.C.; MICHALAK, A.M.; PAHLEVAN, N. Widespread global increase in intense lake phytoplankton blooms since the 1980s. **Nature**, London, v. 574, p. 667–670, 2019.

HOEGER, S.J.; SHAW, G.; HITZFELD, B.C.; DIETRICH, D.R. Occurrence and elimination of cyanobacterial toxins in two Australian drinking water treatment plants. **Toxicon**, Oxford, v. 43, p. 639–649, 2004.

HUMPHREY, J.M.; AGGEN, J.B.; CHAMBERLIN, A.R. Total Synthesis of the Serine-Threonine Phosphatase Inhibitor Microcystin-LA. **Journal of the American Chemical Society**, Washington, DC, v. 118, n. 47, p. 11759-11770, 1996.

JOCHIMSEN, E.M.; CARMICHAEL, W.W.; AN, J.S.; CARDO, D.M.; COOKSON, S.T.; HOLMES, C.E.; ANTUNES, M.B.; DE MELO FILHO, D.A.; LYRA, T.M.; BARRETO, V.S.; AZEVEDO, S.M.; JARVIS, W.R. Liver failure and death after exposure to microcystins at a hemodialysis center in Brazil. **New England Journal of Medicine**, New York, v. 338, n. 13, p. 873–878, 1998.

LIU, Y.; CHEN, W.; LI, D.; SHEN, Y.; LI, G.; LIU, Y. First report of aphantoxins in China - waterblooms of toxigenic *Aphanizomenon flos-aquae* in Lake Dianchi. **Ecotoxicology and Environmental Safety**, Amsterdam, v. 65, p. 84–92, 2006.

MAHMOOD, N.A.; CARMICHAEL, W.W. The pharmacology of anatoxin-a(S), a neurotoxin produced by the freshwater cyanobacterium *Anabaena flos-aquae* NRC 525-17. **Toxicon**, Oxford, v. 24, n. 5, p. 425–434, 1986.

MAHMOOD, N.A.; CARMICHAEL, W.W. Anatoxin-a(S), an anticholinesterase from the cyanobacterium *Anabaena flos-aquae* NRC-525–17. **Toxicon**, Oxford, v. 25, n. 11, p. 1221-1227, 1987.

MAHMOOD, N.A.; CARMICHAEL, W.W.; PFAHLER, D. Anticholinesterase poisoning in dogs from cyanobacterial (blue green algae) bloom dominated by *Anabaena flos-aquae*. **American Journal of Veterinary Research**, Chicago, v. 49, n. 4, p. 500–503, 1998.

MATSUNAGA, S.; MOORE, R.E.; NIEMCZURA, W.P.; CARMICHAEL, W.W. Anatoxin-a(S), a potent anticholinesterase from Anabaena flos-aquae. **Journal of the American Chemical Society**, Washington, DC, v. 111, n. 20, p. 8021–8023, 1989.

MOLICA, R.J.R.; OLIVEIRA, E.J.A.; CARVALHO, P.V.V.C.; COSTA, A.N.S.F.; CUNHA, M.C.C.; MELO, G.L.; AZEVEDO, S.M.F.O. Occurrence of saxitoxins and an anatoxin-a(s)-like anticholinesterase in a Brazilian drinking water supply. **Harmful Algae**, Amsterdam, v. 4, n. 4, p. 743–753, 2005.

MONCIARDINI, P.; BERNASCONI, A.; IORIO, M.; BRUNATI, C.; SOSIO, M.; CAMPOCHIARO, L.; LANDINI, P.; MAFFIOLI, S.I.; DONADIO, S. Antibacterial Aromatic Polyketides Incorporating the Unusual Amino Acid Enduracididine. **Journal of Natural Products**, Washington, DC, v. 82, n. 1, p. 35-44, 2019.

MOORE, B.S.; OHTANI, I.; DE KONING, C.B.; MOORE, R.E.; CARMICHAEL, W.W. Biosynthesis of anatoxin-a(S). Origin of the carbons. **Tetrahedron Letters**, Oxford, v. 33, p. 6595-6598, 1992.

MOORE, R.E.; OHTANI, I.; MOORE, B.S. DE KONING. C.B.; YOSHIDA, W.Y.; RUNNEGAR, M.T.C.; CARMICHAEL, W.W. Cyanobacterial toxins. **Cazzetta Chimica Italiana**, Rome, v. 123, p. 329-336, 1993.

ONGLEY, S.E.; BIAN, X.; NEILAN, B.A.; MÜLLER, R. Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. **Natural Product Reports**, London, v. 30, p. 1121-1138, 2013.

ONODERA, H.; OSHIMA, Y.; HENRIKSEN, P.; YASUMOTO, T. Confirmation of anatoxin-a(S), in the cyanobacterium *Anabaena lemmermannii*, as the cause of bird kills in Danish lakes. **Toxicon**, Oxford, v. 35, n. 11, p. 1645–1648, 1997.

OPPEDIJK, S.F.; MARTIN, N.I.; BREUKINK, E. Hit 'em where it hurts: the growing and structurally diverse family of peptides that target lipid-II. **Biochimica et Biophysica Acta**, Amsterdam, v. 1858, p. 947–957, 2015.

PAERL, H.W.; HUISMAN, J. Climate change: a catalyst for global expansion of harmful cyanobacterial blooms. **Environmental Microbiology Reports**, London, v. 1, p. 27-37, 2009.

PAERL, H.W.; PAUL, V.J. Climate change: links to global expansion of harmful cyanobacteria. **Water Research**, Amsterdam, v. 46, p. 1349-1363, 2012.

PAERL, H.W.; HALL, N.S.; CALANDRINO, E.S. Controlling harmful cyanobacterial blooms in a world experiencing anthropogenic and climatic-induced change. **Science of the Total Environment**, Amsterdam, v. 409, p. 1739–1745, 2011.

PEARSON, L.A.; DITTMANN, E.; MAZMOUZ, R.; ONGLEY, S.E.; D'AGOSTINO, P.M.; NEILAN, B.A. The genetics, biosynthesis and regulation of toxic specialized metabolites of cyanobacteria. **Harmful Algae**, Amsterdam, v. 54, p. 98-111, 2016.

PETRIKOVICS, I.; PAPAHADJOPOULOS, D.; HONG, K.; CHENG, T.C.; BASKIN, S.I.; JIANG, J.; JASZBERENYI, J.C.; LOGUE, B.A.; SZILASI, M.; MCGUINN, W.D.; WAY, J.L. Comparing therapeutic and prophylactic protection against the lethal effect of paraoxon. **Toxicological Science**, San Diego, v. 77, p. 258-262, 2004.

PITA, R.; ANADON, A.; MARTINEZ-LARRANAGA, M.R. Neurotoxins with anticholinesterase activity and their possible use as warfare agents. **Medicina Clinica**, Barcelona, v. 121, p. 511-517, 2003.

RALSTON, D.K.; MOORE, S.K. Modeling harmful algal blooms in a changing climate. **Harmful Algae**, Amsterdam, v. 91, 2020. doi: 10.1016/j.hal.2019.101729.

SIVONEN, K.; JONES, G. Cyanobacterial toxins. In: CHORUS, I.; BARTRAM, J. (Ed.). **Toxin cyanobacteria in water**: a guide to their public health consequences, monitoring and management. London, UK: WHO, 1999. p. 41–111.

STEWART, I.; SEAWRIGHT, A.A.; SHAW, G.R. Cyanobacterial poisoning in livestock, wild mammals and birds - an overview. In**:** HUDNELL, H.K. (Ed.). **Cyanobacterial Harmful Algal Blooms**: State of the Science and Research Needs. New York: Springer, 2008. p. 613–637.

THOMAS, M.G.; CHAN, Y.A.; OZANICK, S.G. Deciphering tuberactinomycin biosynthesis: isolation, sequencing, and annotation of the viomycin biosynthetic gene cluster. **Antimicrobial Agents and Chemotherapy**,Washington, DC, v. 47, n. 9, p. 2823–2830, 2003.

TUOVINEN, K.; KALISTE-KORHONEN, E.; RAUSHEL, F.M.; HÄNNINEN, O. Success of pyridostigmine, physostigmine, eptastigmine and phosphotriesterase treatments in acute sarin intoxication. **Toxicology**, Amsterdam, v. 134, p. 169-178, 1999.

URBAN, L.; HOLZER, A.; BARONAS, J.J.; HALL, M.; BRAEUNINGER-WEIMER, P.; SCHERM, M.J.; KUNZ, D.J.; PERERA, S.N.; MARTIN-HERRANZ, D.E.; TIPPER, E.T.; SALTER, S.J.; STAMMNITZ, M.R. Freshwater monitoring by nanopore sequencing. **bioRxiv**, Huntington, 2020. doi: 10.1101/2020.02.06.936302. Acesso em: Feb. 17, 2020.

VAN APELDOORN, M.E.; VAN EGMOND, H.P.; SPEIJERS, G.J.A.; BAKKER, G.J.I. Toxins of cyanobacteria. **Molecular Nutrition and Food Research**, Weinheim, v. 51, p. 7-60, 2007.

VAN HALDEREN, A.; HARDING, W.R.; WESSELS, J.C.; SCHNEIDER, D.J.; HEINE, E.W.; VAN DER MERWE, J.; FOURIE, J.M. Cyanobacterial (blue-green algae) poisoning of livestock in the western Cape Province of South Africa. **Journal of the South African Veterinary Association**, Cape Town, v. 66, n. 4, p. 260–264, 1995.

WIESE, M.; D'AGOSTINO, P.M.; MIHALI, T.K.; MOFFITT, M.C.; NEILAN, B.A. Neurotoxic Alkaloids: Saxitoxin and Its Analogs. **Marine Drugs**, Basel, v. 8, p. 2185-2211, 2010.

ZHANG, J.J.; TANG, X.; ZHANG, M.; NGUYEN, D.; MOORE, B.S. Broad-host-range expression reveals native and host regulatory elements that influence heterologous antibiotic production in Gram-negative bacteria. **mBio**. Washington, DC, v. 8, n. 5, e01291-17, 2017.

2 GUANITOXIN, RE-NAMING A CYANOBACTERIAL ORGANOPHOSPHATE TOXIN[1]

Abstract

Anatoxin-a(S) is the most potent natural neurotoxin produced by freshwater cyanobacteria. It is also the least understood and monitored. Although this potent cholinesterase inhibitor was first reported in the 1970s and connected with animal poisonings, the lack of chemical standards and identified biosynthetic genes together with limited diagnostics and acute reactivity of this naturally occurring organophosphate have limited our understanding of its environmental breadth and human health implications. Anatoxin-a(S) irreversibly inhibits acetylcholinesterase much like other organophosphate agents like paraoxon. It is however often confused with the similarly named anatoxin-a that has a completely different chemical structure, mechanism of action, and biosynthesis. Herein we propose renaming of anatoxin-a(S) to clarify its distinct structure and mechanism and to draw renewed attention to this potent natural poison. We propose the new name guanitoxin (GNT) to emphasize its distinctive guanidino organophosphate chemical structure.

Keywords: Anatoxin. Cyanobacteria. Harmful algal bloom. Neurotoxin.

In the late 1950's and early 1960's, sporadic outbreaks of lethally poisoned animals were reported to be caused by planktonic *Anabaena* (now called *Dolichospermum*) cyanobacterial blooms in western Canada (Gorham 1962; Gorham et al., 1964). Two single filament strains of *Anabaena flos-aquae,* NRC 44-1 from Burton Lake, Saskatchewan and NRC 525-17 from Buffalo Pound Lake, Saskatchewan, were isolated and form the basis of some of the early foundational work with the *Anabaena* harmful algal bloom toxins (Carmichael and Gorham, 1978). NRC-44-1 cells or culture filtrates intraperitoneally injected at minimal lethal doses in mice (LD_{min} (i.p.m.)) caused death in 2 – 7 min, preceded by signs of paralysis, tremors and mild convulsions (Carmichael and Gorham, 1977). Although the toxin of the strain NRC 525-17 had the same LD_{min} (i.p.m.) and survival time as from strain NRC 44-1, it also produced severe salivation and chromodacryorrhea (bloody tears) in mice, rats, and chicks just prior to respiratory arrest suggesting a second toxin (Carmichael and Gorham, 1978). Consequently,

[1] Fiore, M.F., de Lima, S.T., Carmichael, W.W., McKinnie, S.M.K., Chekan, J.R., Moore, B.S. Guanitoxin, re-naming a cyanobacterial organophosphate toxin. **Harmful Algae**, v. 92, 101737, Feb. 2020. *Article accepted: Dec. 28, 2019. Available online Jan. 14, 2020.

it was proposed at that time that there were two *Anabaena* molecules with similar toxicology profiles, and thus they were provisionally designated anatoxin-a and anatoxin-a(S) in which "S" referred to the salivation phenotype. These authors stated, "final naming of the toxins must await critical studies of toxicology and pharmacology of each type". Therefore, these similar names were given before there was an understanding of whether their chemical structures or molecular mechanism of actions were related.

Over the ensuing four decades, major advances were achieved in understanding the chemistry and biology of the neurotoxic anatoxin-a. Its molecular structure was established as a novel tropane-like alkaloid (Devlin et al., 1977; Huber, 1972; Skulberg et al., 1992) derived biosynthetically from polyketide extension of proline (Méjean et al., 2009). Its mechanism of action was found to be that of a potent depolarizing neuromuscular blocking agent (Carmichael et al., 1975, Carmichael et al., 1979). Additional anatoxin-a congeners were subsequently discovered, namely homoanatoxin-a, dihydroanatoxin-a and dihydrohomoanatoxin-a, as well as numerous cyanobacterial genera producers from worldwide locations (*Dolichospermum* – formerly *Anabaena*, *Kamptonema* – formerly *Oscillatoria*, *Aphanizomenon*, *Cylindrospermum*, *Microcystis*, *Raphidiopsis*, *Planktothrix*, *Nostoc*, *Phormidium*, *Arthrospira*, *Hydrocoleum*, *Cuspidothrix* – formerly *Aphanizomenon issatschenkoi*) (Ballot et al., 2004; Ghassempour et al., 2005; Gugger et al., 2005; Méjean et al., 2010; Namikoshi et al., 2003; Park et al., 1993; Sivonen et al., 1989; Viaggiu et al., 2004; Wood et al., 2007). This toxin and the causative cyanobacterial strains are now regularly monitored worldwide in major fresh water sources to understand and mitigate its occurrence, distribution and effects on human and animal health.

Much less is known about anatoxin-a(S). It was shown to have a fundamentally different mechanism of action as an active site-directed irreversible inhibitor of acetylcholinesterase (EC3.1.1.7) (Mahmood and Carmichael, 1986, 1987; Hyde and Carmichael, 1991) rather than as a nicotinic acetylcholine receptor agonist as in the case of anatoxin-a. Significantly its molecular structure was established as a reactive organophosphate (Matsunaga et al., 1989), and some steps of its biosynthetic pathway were proposed from isotope precursor feeding experiments that showed an origin from the amino acid arginine (Moore et al., 1992; Hemscheidt et al., 1995). However, less attention was given to anatoxin-a(S) for several reasons: fewer validated poisoning events, instability and lack of the purified, standard compound, and the absence of knowledge of its biosynthesis genes. Notwithstanding, this toxin has been reported in North America (Carmichael and Gorham 1978; Mahmood et al., 1988, Cook et al. 1989), South America (Becker et al., 2010; Molica et al., 2005; Monserrat, et al., 2001), Europe (Onodera et al., 1997), and Asia (Chatziefthimiou et al., 2014). Up to now it has

been found only in freshwater planktonic cyanobacterial strains of the genera *Dolichospermum* (previously *Anabaena*) and *Sphaerospermopsis* (Carmichael and Gorham, 1978; Molica et al., 2005).

Traditionally, cyanotoxins have been named according to the first identified producer genus, as in the case of microcystin (*Microcystis*), lyngbyatoxin (*Lyngbya*), cylindrospermopsin (*Cylindrospermopsis*), and anatoxin-a and anatoxin-a(S) (*Anabaena*). The exception is the neurotoxin saxitoxin that was first isolated from a butter clam (*Saxidomus giganticus*), but later discovered to derive from the symbiont marine dinoflagellate *Alexandrium catenella* (formerly *Gonyaulax catenella*) (Sommer et al., 1937) and various cyanobacterial strains. There are some obvious problems with continuing to follow this naming format, one being that cyanobacterial taxonomy is changing ever since cyanobacteria were recognized as a bacterial phylum (Castenholz and Waterbury, 1989; Stanier and Cohen-Bazire, 1977) and classification began to be mainly inferred through molecular phylogeny (Giovannoni et al., 1988; Woese, 1987). For example, currently, only benthic strains of the genus *Anabaena* still retain this name, but the planktonic strains are now renamed as *Dolichospermum* (Komarek, 2013). Therefore, the *Anabaena* strains identified as the first producers of anatoxin-a and anatoxin-a(S) are now reassigned as *Dolichospermum*. Furthermore, the similarity of the terms anatoxin-a and anatoxin-a(S) are confusing and causing difficulty in identifying the right toxin. Therefore, we propose to change the name of the anatoxin-a(S) on the basis of the following reasons:

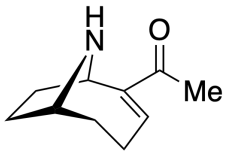
1. The molecular structures of both toxins are fundamentally different (Figure 1). Anatoxin-a is an alkaloid, while anatoxin-a(S) is an organophosphate. The 165 Da anatoxin-a molecule is a bicyclic secondary amine with the general structure 1-((6*R*)-9-azabicyclo[4.2.1]non-2-en-2-yl)ethan-1-one, $C_{10}H_{15}NO$. There are three known analogues, homoanatoxin-a, dihydroanatoxin-a and dihydrohomoanatoxin-a, which differ by a substitution of the acetyl group for a propionyl group or reduction of the C2-C3 olefin (Skulberg et al., 1992). The 252 Da anatoxin-a(S) molecule, on the other hand, is (*S*)-5-((dimethylamino)methyl)-2-iminoimidazolidin-1-yl methyl hydrogen phosphate, $C_7H_{17}N_4O_4P$. No analogues are known.

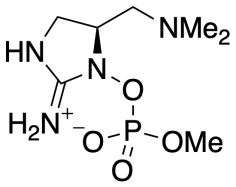
2. Anatoxin molecules are biosynthesized via profoundly different pathways as anticipated from their very different chemical structures. In the case of anatoxin-a, its biosynthesis was firmly established by analysis of its gene cluster from *Oscillatoria* sp. PCC 6506 and reconstitution of the purified enzymes to reveal a modular polyketide synthase

initiated by L-proline (Méjean et al., 2009). The anatoxin-a gene cluster has since been identified and sequenced in several other cyanobacteria suggesting a similar evolutionary origin (Méjean et al., 2014, Ballot et al., 2018). While the anatoxin-a(S) biosynthesis gene cluster has yet to be identified, preliminary isotope tracer experiments have revealed a biosynthetic origin from arginine (Moore et al., 1992; Hemscheidt et al., 1995) that would be independent of carrier protein-based assembly as observed in most other cyanotoxins (Pearson et al., 2016).

3. The toxicity of anatoxin-a(S) is about ten times more potent than anatoxin-a (i.p. LD_{50} in mice of anatoxin-a and anatoxin-a(S) is 200-250 µg/kg of body weight and 20-40 µg/kg, respectively).

4. Although both toxins act in the cholinergic system, the mechanism of toxicity is different between the neurotoxins. Anatoxin-a is a nicotinic acetylcholine receptor agonist that upon binding induces the opening of the receptor channel and depolarization of the cell membrane. Anatoxin-a(S), on the other hand, acts as an irreversible inhibitor of acetylcholinesterase through a proposed covalent bond modification between the serine residue of acetylcholinesterase and the phosphate group of the toxin (Valério et al., 2010; Patocka et al., 2011). Once anatoxin-a(S) binds to acetylcholinesterase, the recycling of acetylcholine is inhibited and there is an over stimulation of the muscles, leading to convulsions, muscle fatigue, and respiratory arrest.





Toxin name:	**anatoxin-a**	**anatoxin-a(S)**
Cyanobacterial producers:	>13 genera	2 genera
Biosynthesis:	**proline + PKS**	**arginine origin**
LD50 in mice (i.p.):	200-250 *µ*g/kg	20-40 *µ*g/kg
Molecular target / MOA:	**Nicotinic acetylcholine receptor agonist**	**Acetylcholinesterase irreversible inhibitor**

Figure 1 – Chemical structures of anatoxin-a and anatoxin-a(S) and related characteristics of both neurotoxins

The literature is rife with articles on anatoxin in which the specific toxin subtype is not clearly articulated but normally refers to anatoxin-a (Draisci et al., 2001; Cadel-Six et al., 2007; Heath et al., 2016). It is thus for all these collective reasons that we recommend a name change and select the lesser known anatoxin-a(S) to be renamed.

While we recognize the merits and difficulties to naming natural compounds, we advocate that the new toxin name should reflect its chemical composition and/or its mechanism of action, which is a more convenient and stable manner of referring to cyanobacterial toxins. The chemical composition of anatoxin-a(S) includes guanidine and phosphate groups and it acts via inhibition of cholinesterase; thus, we propose the name guanitoxin and abbreviation GNT to highlight its core structure of a guanidino group, which accompanies the molecule assembly from the beginning to the end.

References

Ballot, A., Krienitz, L., Kotut, K., Wiegand, C., Metcalf, J.S., Codd, G.A., Pfugmacher, S. 2004. Cyanobacteria and cyanobacterial toxins in three alkaline Rift Valley lakes of Kenya-lakes Bogoria, Nakuru and Elmenteita. J. Plankton Res. 26:925–935.

Ballot, A., Scherer, P.I., Wood, S.A. 2018. Variability in the anatoxin gene clusters of *Cuspidothrix issatschenkoi* from Germany, New Zealand, China and Japan. PLoS One. 13:e020074.

Becker, V., Ihara, P., Yunes, J.S., Huszar, V.L.M. 2010. Occurrence of anatoxin-a(S) during a bloom of *Anabaena crassa* in a water-supply reservoir in southern Brazil. J. Appl. Phycol. 22:235–241.

Cadel-Six, S., Peyraund-Thomaz, C., Brient, L., Marsac, N.T., Rippka, R., Méjean, A. Different genotypes of anatoxin-producing cyanobacteria coexist in the Tarn River, France. 2007. Appl. Environ. Microbiol. 73: 7605-7614.

Carmichael, W.W., Biggs, D.F., Gorham, P.R. 1975. The toxicology and pharmacological action of *Anabaena flos-aquae* toxin. Science 187:542-544.

Carmichael, W.W., Gorham, P.R. 1977. Factors influencing the toxicity and animal susceptibility of *Anabaena flos-aquae* (Cyanophyta) blooms. J. Phycol. 13:97-101.

Carmichael, W.W., Gorham, P.R. 1978. Anatoxins from clones of *Anabaena flos-aquae* isolated from lakes of western Canada. Mitt. Int.Ver. Theor. Angew. Limnol. 21:285-295.

Carmichael, W.W., Biggs, D.F., Peterson, M.A. 1979. Pharmacology of anatoxin-a: produced by the freshwater cyanophyte *Anabaena flos-aquae* NRC-44-1. Toxicon. 17:229-236.

Castenholz, R.W., Waterbury, J.B. 1989. Oxygenic photosynthetic bacteria: Group I. Cyanobacteria. In: J.T. Staley, M.P. Bryant, N. Pfennig, J.G. Holt (Eds) Bergey's manual of systematic bacteriology. Vol. 3, Williams & Wilkins, Baltimore, 1710-1727.

Chatziefthimiou, A.D., Richer, R., Rowles, H., Powell, J.T., Metcalf, J. 2014. Cyanotoxins as a potential cause of dog poisonings in desert environments. Vet. Rec. 174:484–485.

Cook, W.O., Beasley, V.R., Lovell, R.A., Dahlem, A.M., Hooser, S.B., Mahmood, N.A., Carmichael, W.W. 1989. Consistent inhibition of peripheral cholinesterases by neurotoxins from the freshwater cyanobacterium *Anabaena flos-aquae*: studies of ducks, swine, mice and a steer. Environ. Toxicol. Chem. 8:915-922.

Devlin, J.P., Edwards, O.E., Gorham, P.R., Hunter, N.R., Pike, R.K., Stavric, B. 1977. Anatoxin-a, a toxic alkaloid from *Anabaena-flos-aquae* NRC-44h. Can. J. Chem. 55:1367-1371.

Draisci, R., Ferretti, E., Palleschi, L., Marchiafava, C. 2001. Identification of anatoxins in blue-green algae food supplements using liquid chromatography-tandem mass spectrometry. Food Addit. Contam. 18: 525-531.

Ghassempour, A., Najafi, N.M., Mehdinia, A., Davarani, S.S.H., Fallahi, M., Nakhshab, M. 2005. Analysis of anatoxin-a using polyaniline as a sorbent in solid-phase microextraction coupled to gas chromatography-mass spectrometry. J. Chromatogr. A 1078:120–127.

Giovannoni, S.J., Turner, S., Olsen, G.J., Barns, S., Lane, D.J., Pace, N.R. 1988. Evolutionary relationships among cyanobacteria and green chloroplasts. J. Bacteriol. 170:3584-3592.

Gorham, R.P. 1962. Laboratory studies on the toxins produced by waterblooms of blue-green algae. Am. J. Public Health Nations Health 52:2100-21005.

Gorham, R.P., McLachlan, J., Hammer, U.T., Kim, W.K. 1964. Isolation and culture of toxic strains of *Anabaena flos-aquae* (Lyngb.) de Bréb. Verh. Internat. Verein. Limnol. 15:796-804.

Gugger, M., Lenoir, S., Berger, C., Ledreux, A., Druart, J.-C., Humbert, J.-F., Guette, C., Bernard, C. 2005. First report in a river in France of the benthic cyanobacterium *Phormidium favosum* producing anatoxin-a associated with dog neurotoxicosis. Toxicon 45:919–928.

Heath, M., Wood, S.A., Young, R.G., Ryan, K.G. 2016. The role of nitrogen and phosphorus in regulating *Phormidium* sp. (cyanobacteria) growth and anatoxin production. FEMS Microbiol. Ecol. 92: 1-8.

Hemscheidt, T., Burgoyne, D.L., Moore, R.E. 1995. Biosynthesis of anatoxin-a(S). (2*S*,4*S*)-4-hydroxyarginine as an intermediate. J. Chem. Soc., Chem. Commun. 2:205-206.

Hyde, E.G., Carmichael, W.W. 1991. Anatoxin-a(S), a naturally occurring organophosphate, is an irreversible active site-directed inhibitor of acetylcholinesterase (EC3.1.1.7). J. Biochem. Toxicol. 6:195-201.

Huber, C.S. 1972. The crystal structure and absolute configuration of 2,9-diacetyl-9-azabicyclo[4,2,1]non-2,3-ene. Acta Crystallograph. B28:2577-2582.

Komarek, J. 2013. Suβwasserflora von Mitteleuropa. Bd. 19/3: Cyanoprokaryonta. Springer-Verlag Berlin Heidelberg, 1130 pg.

Mahmood, N.A., Carmichael, W.W. 1986. The pharmacology of anatoxin-a(S), a neurotoxin produced by the freshwater cyanobacterium *Anabaena flos-aquae* NRC 525-17. Toxicon 24:425-434.

Mahmood, N.A., Carmichael, W.W. 1987. Anatoxin-a(S), an irreversible anticholinesterase from the cyanobacterium *Anabaena flos-aquae* NRC-525-17. Toxicon 25: 1221-1227.

Mahmood, N.A., Carmichael, W.W., Pfahler, D. 1988. Anticholinesterase poisonings in dogs from a cyanobacterial (blue-green-algae) bloom dominated by *Anabaena flos-aquae*. Am. J. Vet. Res. 49:500–503.

Matsunaga, S., Moore, R.E., Niemczura, W.P., Carmichael, W.W. 1989. Anatoxin-a(S), a potent anticholinesterase from *Anabaena flos-aquae*. J. Am. Chem. Soc. 111:8021-8023.

Méjean, A., Mann, S., Maldiney, T., Vassiliadis, G., Lequin, O., Ploux, O. 2009. Evidence that biosynthesis of the neurotoxic alkaloids anatoxin-a and homoanatoxin-a in the cyanobacterium *Oscillatoria* PCC 6506 occurs on a modular polyketide synthase initiated by L-proline. J. Am. Chem. Soc. 131:7512-7513.

Méjean, A., Peyraud-Thomas, C., Kerbrat, A.S., Golubic, S., Pauillac, S., M. Chinain, M., Laurent, D. 2010. First identification of the neurotoxin homoanatoxin-a from mats of *Hydrocoleum lyngbyaceum* (marine cyanobacterium) possibly linked to giant clam poisoning in New Caledonia. Toxicon 56:829–835.

Méjean, A., Paci, G., Gautier, V., Ploux, O. 2014. Biosynthesis of anatoxin-a and analogues (anatoxins) in cyanobacteria. Toxicon. 91:15-22.

Molica RJR, Oliveira EJA, Carvalho PVVC, Costa ANSF, Cunha MCC, Melo GL, Azevedo SMFO. 2005. Occurrence of saxitoxins and an anatoxin-a(S)-like anticholinesterase in a Brazilian drinking water supply. Harmful Algae, 4:743–753.

Monserrat, J.M., Yunes, J.S., Bianchini, A. 2001. Effects of *Anabaena spiroides* (Cyanobacteria) aqueous extracts on the acetylcholinesterase activity of aquatic species. Environ. Toxicol. Chem. 20:1228-1235.

Moore, B.S., Ohtani, I., de Koning, C.B., Moore, R.E., Carmichael, W.W. 1992. Biosynthesis of anatoxin-a(S). Origin of the carbons. Tetrahedron Lett. 33:6595–6598.

Namikoshi, M., Murakami, T., Watanabe, M.F., Oda, T., Yamada, J., Tsujimura, S., Nagai H., Oishi, S. 2003. Simultaneous production of homoanatoxin-a, anatoxin-a, and a new non-toxic 4-hydroxyhomoanatoxin-a by the cyanobacterium *Raphidiopsis mediterranea* Skuja. Toxicon 42:533–538.

Onodera, H., Oshima, Y., Henriksen, P., Yasumoto, T. 1997. Confirmation of anatoxin-a(S), in the cyanobacterium *Anabaena lemmermannii*, as the cause of bird kills in Danish lakes. Toxicon 35:1645–1648.

Park, H.D., Watanabe, M.F., Harada, K.-I., Nagai, H., Suzuki, M., Watanabe, M., Hayashi, H. 1993. Hepatotoxin (microcystin) and neurotoxin (anatoxin-a) contained in natural blooms and strains of cyanobacteria from Japanese freshwaters. Nat. Toxins. 1:353–360.

Patocka, J., Gupta, R.C., Kuca, K. 2011. Anatoxin-a(S): natural organophosphorus anticholinesterase agente. Mil. Med. Sci. Lett. 80: 129-139.

Pearson, L.A., Dittmann, E., Mazmouz, R., Ongley, S.E., D'Agostino, P.M., Neilan, B.A. 2016. The genetics, biosynthesis and regulation of toxic specialized metabolites of cyanobacteria. Harmful Algae. 54:98-111.

Sivonen, K., Himberg, K., Luukkainen, R., Niemelă, S.I., Poon, G.K., Codd, G.A. 1989. Preliminary characterization of neurotoxic cyanobacteria blooms and strains from Finland. Environ. Toxicol. 4:339–352.

Skulberg, O.M., Carmichael, W.W., Andersen, R.A., Matsunaga, S., Moore, R.E., Skulberg, R. 1992. Investigations of a neurotoxic Oscillatorialean strain (Cyanophyceae) and its toxin. Isolation and characterization of homoanatoxin-a. Environ. Toxicol. Chem. 11:321-329.

Sommer, H., Whedon, W.F., Kofoid, C.A., Stohler, R. 1937. Relation of paralytic shell-fish poison to certain plankton organisms of the genus *Gonyaulax*. Arch. Pathol. 37:537-559.

Stanier, R.Y., Cohen-Bazire, G. 1977. Photosynthetic prokaryotes: the cyanobacteria. Ann. Rev. Microbiol. 31:225-275.

Valério, E., Chaves, S., Tenreiro, R. 2010. Diversity and impact of prokaryotic toxins on aquatic environments: a review. Toxins. 2(10): 2359-2410.

Viaggiu, E., Melchiorre, S., Volpi, F., Di Corcia, A., Mancini, R., Garibaldi, L., Crichigno, G., Bruno, M. 2004. Anatoxin-a toxin in the cyanobacterium *Planktothrix rubescens* from fishing pond in northern Italy. Environ. Toxicol. 19:191–197.

Woese C. 1987. Bacterial evolution. Microbiol. Rev. 51:221–271.

Wood, S.A., Rasmussen, J.P., Holland, P.T., Campbell, R., Crowe, A.L.M. 2007. First report of the cyanotoxin anatoxin-a from *Aphanizomenon issatschenkoi* (Cyanobacteria). J. Phycol. 43:356–365.

3 DISENTANGLING THE BIOSYNTHESIS OF A NEUROTOXIC NATURAL ORGANOPHOSPHATE IN CYANOBACTERIA[2]

Abstract

Toxic freshwater cyanobacterial harmful algae blooms (CyanoHBs) are a public health concern worldwide. One of the most potent and overlooked cyanotoxins, guanitoxin (GNT), had its chemical structure elucidated in the late 1980's. However, the genes encoding its biosynthetic pathway are still unknown. By using genome mining approaches, we found a potential biosynthetic gene cluster (BGC) responsible for GNT production. We then biochemically investigated the GNT production using *in vitro* and *in vivo* assays to rigorously validate our proposal. Using our results, we were able to show that the GNT BGC was actively expressed in one of the largest cyanobacterial blooms hotspots, Lake Erie, Toledo, USA. Our GNT biosynthetic proposal may contribute to the development of rapid and sensitive methods of detection and monitoring of this forgotten toxin in freshwaters used for public supply.

There has been rise of cyanobacterial harmful algal blooms (CyanoHBs) worldwide and they remain significant threat to the health of humans and animals (BURFORD et al., 2020). Notorious accidents include the fatal cyanotoxin intoxication at a hemodialysis unit in Caruaru (Pernambuco, Brazil), which led to more than 60 deaths (JOCHIMSEN et al., 1998) and the Toledo (Ohio, USA) water crisis in 2014 that left nearly a half million people without potable water for two days because a toxic cyanobacterial bloom overwhelmed Toledo's water treatment facility (BULLERJAHN et al., 2016). Additionally, poisonings have been reported for dogs, ducks, swine, sheeps and birds (BACKER et al., 2013; COOK et al., 1989a; FOSS et al., 2019; GUGGER et al., 2005; GORHAM, 1962; HENRIKSEN et al., 1997; MAHMOOD et al., 1988; METCALF et al., 2006; NEGRI et al., 1995). Several bloom-forming cyanobacteria are producers of potent toxins that include microcystin, nodularin, cylindrospermopsin, saxitoxin and anatoxin-a. The biosynthetic pathways for these toxins were already elucidated, allowing studies of their distribution, toxicity and monitoring. However, one of the most toxic cyanotoxins known, guanitoxin (GNT, formerly known as anatoxin-a(S), FIORE et al., 2020) remains poorly documented. GNT has been reported in the North America, South America,

[2] Lima, S.T., Chekan, J.R., McKinnie, S.M.K., Alvarenga, D.O., Luhavaya, H., Dörr, F.A., Etchegaray, A., Pinto, E., Renata, H., Moore, B.S., Fiore, M.F. Disentangling the biosynthesis of a neurotoxic natural organophosphate in cyanobacteria. Manuscript in preparation.

(CARMICHAEL; GORHAM, 1978; MAHMOOD et al., 1988; MOLICA et al., 2005; MONSERRAT et al., 2001; BECKER et al., 2010), Europe (ONODERA et al., 1997) and Asia (CHATZIEFTHIMIOU et al., 2014) continents. Fresh water planktic strains of the genera *Dolichospermum* and *Sphaerospermopsis* (Fig. S2) have been known as GNT producers. The occurrence of GNT is likely to remain underestimated until major advances in genetic and biochemical route for its production by cyanobacteria will appear.

The toxicity of GNT is based on its irreversible inhibition of acetylcholinesterase (HYDE; CARMICHAEL, 1991; MAHMOOD; CARMICHAEL, 1986, 1987). The anti-cholinesterase activity of GNT is effectively exerted at the peripheral nervous system, while the brain and retinal acetylcholinesterase activities remain normal, even in lethally poisoned animals (COOK et al., 1988; COOK et al., 1989b). The affinity for the human acetylcholinesterase in red blood cells is relatively high, therefore the risk of acute intoxication for humans is higher than for some aquatic species (CARMICHAEL et al., 1990). The acute neurological effects in mammals are muscle weakness, hypersalivation, respiratory distress (dyspnea), and convulsions preceding death, which occurs due to respiratory arrest (COOK et al., 1988; MAHMOOD; CARMICHAEL, 1987). The LD_{50} in mice after i.p. injection is 20-40 μg/kg of body weight (MAHMOOD; CARMICHAEL, 1987) and no toxicological data are available for deriving an acute dose without effects. In addition, due to the lack of sub-chronic and chronic data, no tolerable daily intake (TDI) or guideline value can yet be derived for GNT. Hazards linked to cyanobacterial contamination have been recognized and addressed by regulatory authorities, and for cyanobacterial bloom management the precautionary principle was proposed, which means that the bloom is considered hazardous until proven safe (CHORUS, 2012; OTTEN; PAERL, 2015).

Four decades ago, the unique chemical structure of GNT was unveiled by the research group of Prof. Richard Moore at the University of Hawaii as the reactive organophosphate (*S*)-5-((dimethylamino)methyl)-2-iminoimidazolidin-1-yl methyl hydrogen phosphate ($C_7H_{17}N_4O_4P$) (MATSUNAGA et al., 1989). Subsequently, some steps of its biosynthetic pathway were proposed from radiolabeled precursor-feeding experiments confirming that arginine is the starter unit and indicating that 4-hydroxy-L-arginine and enduracididine were biosynthetic intermediates (HEMSCHEIDT et al., 1995; MOORE et al., 1992; MOORE et al., 1993). These steps of GNT biosynthesis anticipated a route distinct from the carrier protein-based assembly, observed in most other cyanotoxin biosynthetic pathways (PEARSON et al., 2016). These initial findings paved the way to successfully disentangle the complete synthesis of GNT. Based on the argument that L-arginine was the GNT precursor and on its proposed

partial biosynthesis, a genome-mining approach was applied to the search for the putative GNT biosynthetic gene cluster in the producer strain *Sphaerospermopis torques-reginae* ITEP-024.

To identify the biosynthetic gene cluster, we focused on the key intermediate enduracididine. This cyclized arginine residue is present in several bioactive natural products and had only a single known biosynthetic route. Therefore, we created a database of enzymes involved in the biosynthesis of enduracididine containing natural products mannopeptimycin (HAN et al., 2018), viomycin (HELMETAG et al., 2009), and enduracididine (ATKINSON et al., 2016) and searched for their presence in the *S. torques-reginae* ITEP-024 genome. Using this approach, a homolog to the PLP dependent oxygenase MppP was found. This enzyme catalyzes the first committed step of known enduracididine biosynthesis to generate a 4-hydroxy-alpha keto arginine derivative. This gene, *gntC*, was found in a candidate metabolic gene cluster consisting of 10 co-localized genes (*gntA-J*) (Fig. 1C).

A closer look at the putative GNT gene cluster revealed that contained genes that could be used to construct a hypothetical biosynthetic pathway. GntC and GntD showed high degree of amino acid sequence similarity with the guanidino skeleton-forming enzymes OrfR and OrfP from streptolidine biosynthesis. OrfP is a non-heme αKG-dependent dihydroxylase that produces the substrate for OrfR, is a PLP-dependent aminotransferase that catalyzes cyclization of the guanidino containing 6-membered ring (CHANG et al., 2014). Furthermore, GntD also has significant amino acid sequence similarity with MppO from mannopeptimycin biosynthesis, an β-enduracididine hydroxylase (HALTLI et al., 2005). In addition to similarity to the arginine cyclization enzymes, the hypothetical GNT pathway also contains another enzyme predicted to use an arginine like substrate, GntA. GntA shares amino acid sequence similarity with YqcI/YcgG family proteins, which include a *N*(ω)-hydroxy-L-arginine synthase (DcsA) from the D-cycloserine (DCS) biosynthetic gene cluster. The heme protein, DcsA, is responsible for the first hydroxylation of L-arginine in the predicted pathway of hydroxyurea formation in DCS biosynthesis. However, this enzymatic reaction was not confirmed by *in vitro* assays (KUMAGAI et al., 2012; KUMAGAI et al., 2015). Remaining enzymes in the GNT gene cluster include an aldolase, kinase and methyltransferases, which have enzymatic activities that are closely related to the chemical steps proposed by Moore and collaborators (MOORE et al., 1992; MOORE et al., 1993; HEMSCHEIDT et al., 1995). Thus, previous findings linked with genome-mining and predicted GNT enzyme functions, enabled us to connect these genes to the guanitoxin biosynthesis.

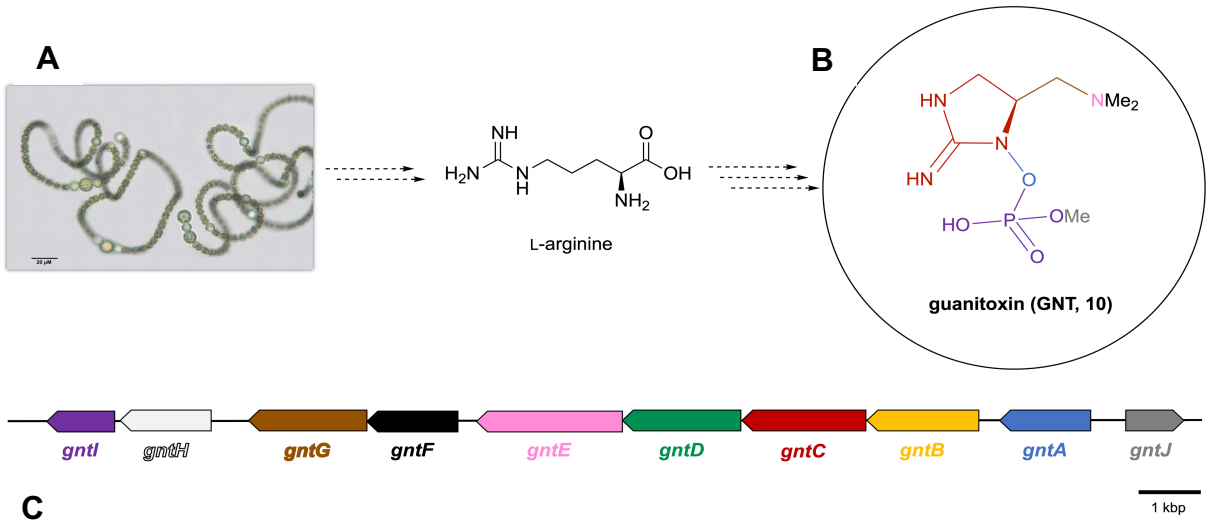


Fig. 1 – Identification of guanitoxin biosynthetic gene cluster from the *Sphaerospermopsis torques-reginae* ITEP-024 genome. **(A)** *Sphaerospermopsis torques-reginae* ITEP-024 culture image, adapted from WERNER et al., 2012. **(B)** Chemical structure of GNT. **(C)** GNT biosynthetic gene cluster (*gnt*) organization

Table 1 - Deduced functions for proteins in the guanitoxin biosynthetic gene cluster

Protein	Amino Acids	Proposed Function	Protein Homolog and Origin	Identities	Accession number
GntA	267	N-hydroxylase	YqcI/YcgG family protein [*Calothrix*]	77%	WP_096685072.1
GntB	318	Arginine hydroxylase	Fatty acid desaturase [*Calothrix* sp. NIES-2098]	52%	WP_096589832.1
GntC	370	hydroxyarginine PLP-dependent cyclase	Aminotransferase class I/II-fold pyridoxal phosphate-dependent enzyme [*Calothrix* sp. NIES-2098]	62%	WP_096589833.1
GntD	347	Enduracididine β-hydroxylase	arginine β-hydroxylase Fe(II)/α-ketoglutarate dependent enzyme [*Calothrix desertica*]	45%	WP_127087020.1
GntE	436	Aminotransferase PLP-dependent	Aminotransferase class III-fold pyridoxal phosphate-dependent enzyme [*gamma proteobacterium BW-2*]	80%	WP_149684313.1
GntF	274	N-methyltransferase SAM-dependent	N-methyltransferase [*Leptolyngbya* sp. PCC 7375]	30%	WP_006513898.1
GntG	344	Aldolase	Low-specificity L-threonine aldolase [*Ardenticatena maritima*]	63%	WP_060687126.1

GntH	413	Phosphatase	MBL fold metallo-hydrolase [*Mesorhizobium* sp. F7]	45%	WP_052224966.1
GntI	249	Kinase	Choline/Ethanolamine kinase [*Chloroflexi bacterium OLB14*]	31%	KXK14598.1
GntJ	249	O-methyltransferase SAM-dependent	SAM-dependent methyltransferase [*Chloroflexi bacterium*]	44%	PZC47808.1
GntT	452	Transporter	MATE family efflux transporter [*Scytonema hofmannii* PCC 7110]	57%	KYC36324.1

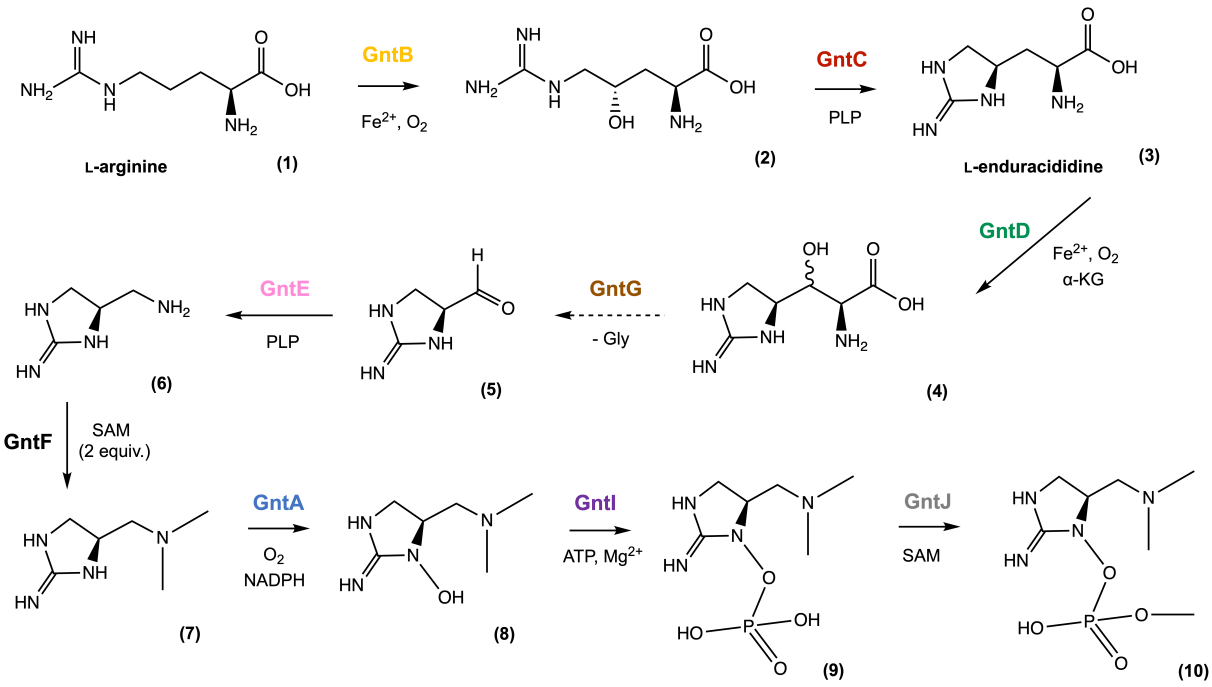


Fig. 2 – Guanitoxin (10) biosynthetic pathway based on GNT biosynthetic gene cluster annotation and enzymatic activities

The validation of the proposed GNT biosynthetic pathway was carried out by *in vivo* and *in vitro* studies. As GntC is most similar to OrfR, which that uses a hydroxylated L-Arg substrate to facilitate cyclization, we hypothesized that the GNT pathway should be initiated with hydroxylation of L-Arg. Bioinformatic analysis predicted that the GntB enzyme is a fatty acid desaturase/oxidoreductase membrane protein, which could conceivably catalyze a hydroxylation event. Unfortunately, GntB proved difficult to express and purify. Therefore, the first (GntB) and second (GntC) enzymes of the GNT route had their gene sequences assembled into the expression vector pCOLADuet^{TM}-1 (Fig. S4) and successfully expressed in *Escherichia coli*. *In vivo*, the recombinant *E. coli* was able to produce a product

with mass consistent with L-enduracididine (**3**) (Fig. S5), the third intermediate of GNT biosynthetic pathway. We hypothesize that this proceeds by an initial hydroxylation of L-Arg at the gamma position by GntB. GntC then performs a cyclodehydration reaction to generate enduracididine (Fig. 2).

We next expressed codon optimized recombinant His_6-GntD in *E. coli* and purified the enzyme using Ni^{2+} affinity chromatography and molecular exclusion chromatography (Fig. S6). *In vitro*, the soluble GntD reacted with pCOLADuet-1_GntBC expression extract in the presence of α-ketoglutarate cofactor, Fe^{2+} and L-ascorbic acid, converting L-enduracididine (**3**) into a product mass consistent with β-hydroxy-L-enduracididine (**4**) by the addition of the hydroxyl group at β-carbon position (Fig. S7).

We next used crude GntBC enduracididine containing production media and examined the proposed biosynthetic enzymes to validate they could progress the enduracididine to (**6**), of which we synthesized a synthetic standard. We used one-pot coupled reactions starting with pCOLADuet-1_GntBC expressed media by sequentially adding purified GntD β-hydroxylase, GntG aldolase, and GntE aminotransferase. In the presence of PLP and L-glutamate the chromatograms showed stepwise progression of enduracididine to a product with the same retention time and extracted mass as synthetic (**6**) (Fig. 3 and Fig. S8). Aldehyde intermediate (**5**) was not able to be detected by LC-MS analysis likely due to its instability in solution (Fig. S8, dark red trace).

The recombinant GntF purified from *E. coli* was able to catalyze the synthesis of (**7**) by adding two S-adenosylmethionine (SAM) equivalents into (**6**) (Fig. 3 and Fig. S9). For the subsequent reactions, GntA was interrogated in the assay the synthetized dimethylamine compound, intermediate (**7**). GntA successfully converted (**7**) into (**8**) to form the *N*-hydroxylated intermediate compound in a nicotinamide adenine dinucleotide phosphate (NADPH) coenzyme-dependent manner (Fig. 3 and Fig. S10).

Structural predictions using Phyre2 and peptide sequence search for similarity detection against a protein database suggested low similarity to any characterized protein in the Protein Data Bank. These results led us to conclude that this enzymatic reaction is a major step in the GNT biosynthetic pathway. Analyzing the remaining steps of the GNT assembly, we deduced that almost all reactions are common to other biosynthetic pathways in the biosynthesis of intermediates used in different classes of natural products (BURROUGHS et al., 2013; HALTLI et al., 2005; HAN et al., 2015; THOMAS et al., 2003). However, none of these include the enzymatic activity of GntA, which is unique to the synthesis of GNT. Thus, the heme protein

N-hydroxylase GntA may be a key enzyme in GNT biosynthesis and, therefore, a hallmark for environmental detection of this toxin.

To continue reconstitution of the biosynthetic pathway, the reaction product of GntA was filtered at molecular weight cut-off filters and recombinant GntI kinase catalyzed the reaction forming the phosphorylated product (**9**) in the presence of Mg^{2+} and NaCl (Fig. 3 and Fig. S11 A). This phosphorylated compound did not appear to ionize very well, consequently making it difficult to detect it by LC-MS analysis. Even though the intensity of expected product was not very high, nearly all the substrate (**8**) was quickly consumed by GntI, indicating high level of kinase activity (Fig. S11 A, blue trace). Finally, the GntJ *O*-methyltransferase and SAM were added to the filtered GntI reaction. After incubation, we achieved the *in vitro* production of the cyanobacterial neurotoxin guanitoxin (**10**) (Fig. 3 and Fig. S11 B).

Overall, a combination of our *in vivo* and *in vitro* biochemical results showed that the GNT biosynthetic pathway that starts with L-arginine and goes until (**6**), passing through the well-known L-enduracididine and β-hydroxy-L-enduracididine in the five reactions to produce the stable intermediate (**6**). Using synthetic (**6**), we then demonstrated that four additional enzymes advances (**6**) to guanitoxin (**10**) (Fig. 2). Thus, these results corroborate with previous studies of radiolabeled intermediates (HEMSCHEIDT et al., 1995; MOORE et al., 1992; MOORE et al., 1993). According to enzymatic action, GntB, a membrane protein, catalyzes the first reaction accepting L-arginine (**1**) as substrate and hydroxylating it at γ-carbon position, producing 4-hydroxy-L-arginine (**2**). GntC takes (**2**) and catalyzes a PLP-dependent cyclization reaction with 4-hydroxy-L-arginine to produce L-enduracididine (**3**). GntD acts to hydroxylate the enduracididine substrate at β-carbon position, producing β-hydroxy-enduracididine (**4**). GntG accepts (**4**) and removes a glycine, producing an aldehyde intermediate (**5**). GntE catalyzes a transamination to produce (**6**) and GntF introduces two methyl equivalents in the nitrogen to produce (**7**). This intermediate hydroxylated by GntA to generate the (**8**) and then receives a phosphate group by the action of GntI to produce (**9**). Finally, (**9**) receives a methyl group by GntJ (SAM-dependent *O*-methyltransferase) to form the final product, guanitoxin (**10**) (Fig. 2). To further validate this biosynthetic proposal, when investigated the presence of GNT BGC intermediates in the *S. torques-reginae* ITEP-024 growth culture, we were able to observe compounds with the same extract mass and retention time as (**7**) and (**8**) (Fig. S12).

GNT has previously been demonstrated to be an acetylcholinesterase inhibitor, which is the basis for its toxicity. With the newly discovered on pathway intermediates, we sought to confirm this activity and determine which biosynthetic intermediates are bioactive. Therefore, acetylcholinesterase inhibition assays were used to investigate the relative activity of (**8**), (**9**) and (**10**). Interestingly, we found that only (10) was bioactive. Therefore, the last biosynthetic step, addition of the methyl group at phosphate group oxygen, is crucial for robust acetylcholinesterase inhibition (Fig. S13).

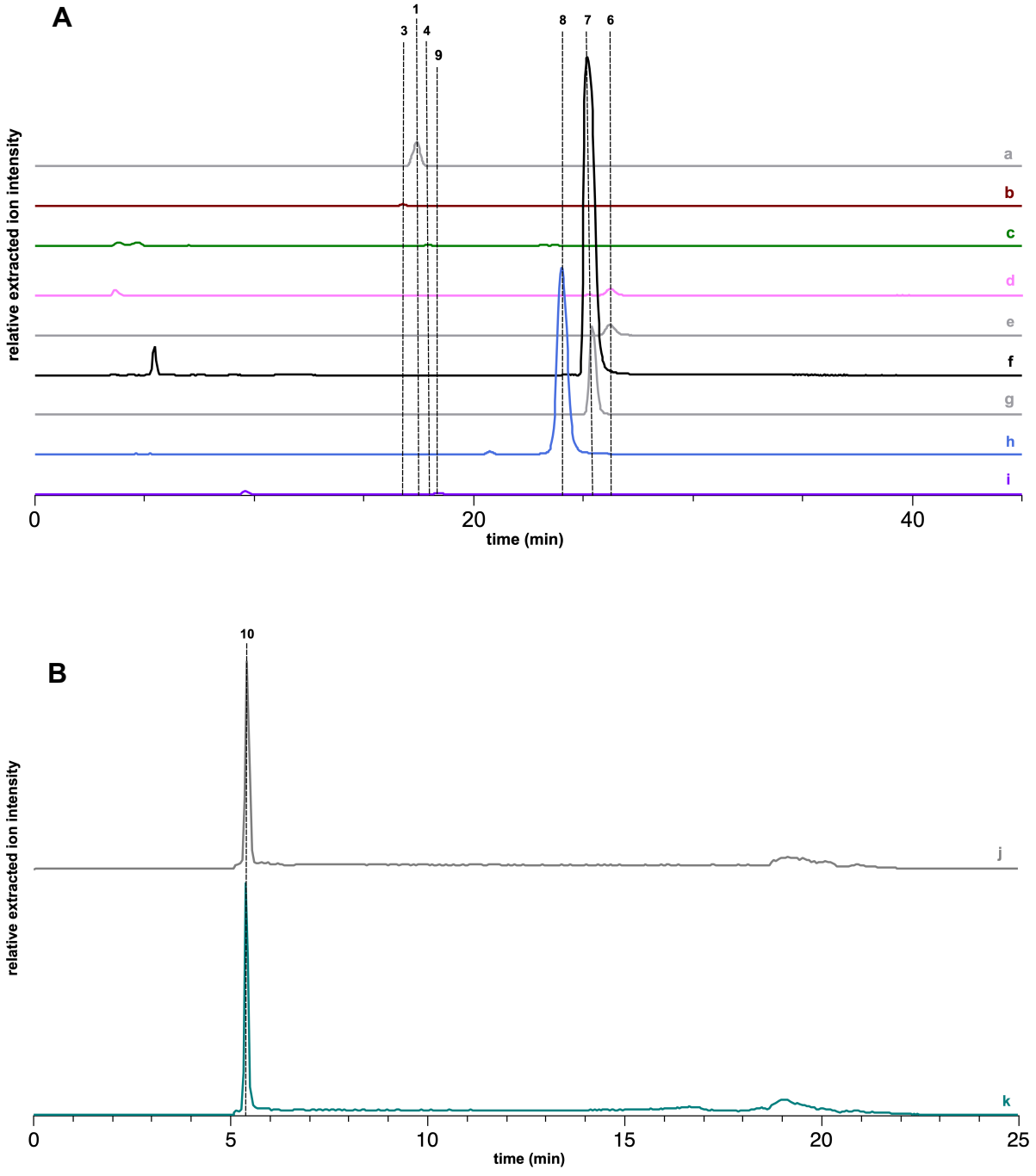


Fig. 3 – *In vivo* and *in vitro* characterization of Gnt enzymatic reactions. (A) Relative intensities of positive ionization – extracted chromatograms Method A liquid chromatography and mass spectrometry dashed traces [(175.1190, 191.1139, 173.1033, 189.0982, 115.0978, 143.1291, 159.1240, 239.0904) ± 0.01 *m/z*] for *in vitro* and *in vivo* Gnt enzymatic assays. a: L-Arg (1: 0.1 mM), Tris buffer pH 7.4 (50 mM); b: pCOLADuet_GntBC expressed for 5 days at 18 ºC; c: pCOLADuet_GntBC extracted, GntD, αKG, L-ascorbic acid, $FeSO_4$; d: pCOLADuet_GntBC extracted, GntD, GntG, GntE, αKG, L-ascorbic acid, $FeSO_4$, PLP, L-Glu; e: synthetic standard (6: 10 μM), Tris buffer pH 7.4 (50 mM); f: GntF, 6 (0.1 mM) standard, SAM, Tris buffer pH 7.4; g: synthetic standard (7: 0.1 mM), Tris buffer pH 7.4; h: GntA, 7 (0.1 mM), NADPH, Tris buffer pH 7.4; i: GntA, GntI, 7 (0.1 mM), NADPH, ATP, NaCl, $MgCl_2$, Tris buffer pH 7.4. (B) Relatives intensities of positive ionization – extracted chromatograms Method B liquid chromatography and mass spectrometry dashed traces (253.1060 ± 0.01 *m/z*). j: GntA, GntI, GntJ, 7 (0.1 mM), NADPH, ATP, NaCl, $MgCl_2$, SAM, Tris buffer pH 7.4; k: *S. torques-reginae* ITEP-024 extracted. Additional enzymatic conditions, controls, cosubstrates and cofactors concentrations can be found in the supplementary materials

BLAST analysis of the National Center for Biotechnology Information (NCBI) database was unsuccessful to reveal any accessible cyanobacterial genome containing GNT biosynthetic genes. Instead, we turned to freshwater metagenomic and metatranscriptomic data sets to look for the occurrence of the GNT biosynthetic genes in the environment. Using this approach, the GNT biosynthetic gene cluster was recovered from metatranscriptomic data of Lake Erie, Toledo, Ohio, USA. The environmental samples of this data set were collected in 2013, in a large cyanobacterial bloom of *Microcystis*, *Anabaena* and *Planktothrix* (HARKE et al., 2016). Assembly of the transcriptome revealed a large polycistronic read that had the same gene organization as the GNT gene cluster from the *S. torques-reginae* ITEP-024 genome (Fig. 4). The genes *gntH* (encoding a phosphatase) and *gntI* (encoding a kinase) (Fig. 4, in purple and orange, respectively), could not be assembled in the Lake Erie data. These two genes were expected to be found in the end of the transcript, where inefficient RNA transcription may have led to low coverage in the transcriptome. However, individual sequencing reads were found to match these genes with high similarity, showing that these genes are indeed present in the transcriptome, but in low abundance. These results suggest the presence of this potent neurotoxin in an aquatic ecosystem used as a major drinking water supplier. Importantly, there are no current programs in place to test for or monitor GNT by health and environmental authorities.

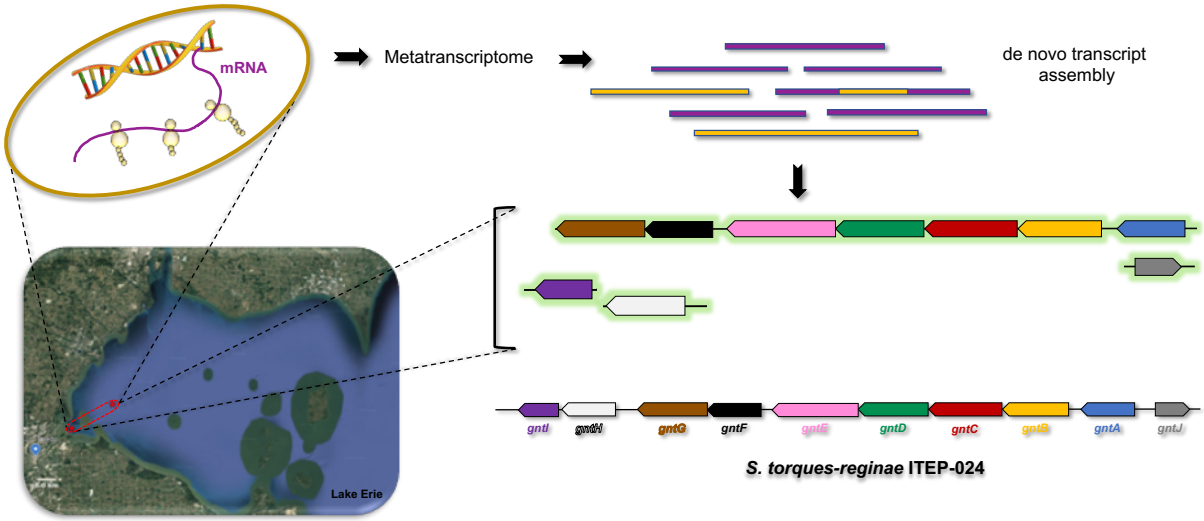


Fig. 4 – GNT biosynthetic gene cluster in Lake Erie metatranscriptomic data from 2013 toxic cyanobacterial bloom

The GNT biosynthetic pathway has a route that is distinct from the other cyanotoxin since it is not assembled by multifunctional non-ribosomal/polyketide biosynthetic enzymes. On the contrary, it uses the amino acid L-arginine as starter unit, and it is elaborated by a set of stand along enzymes all encoded by co-localized genes in the cyanobacterial genomes. We anticipate that the discovery of *gnt* biosynthetic genes will lead to the understanding of toxin distribution and its role in metabolic processes and may help to comprehend the microbial ecology dynamics associated with freshwater blooms. Moreover, the knowledge of the GNT biosynthetic route will contribute to the development of fast and specific detection methods for one of the most potent neurotoxins produced by cyanobacteria. In addition, it may contribute to the establishment of guidelines for short-term and recreational exposure by public health authorities in fresh water supplies.

References

ATKINSON, D.J.; NAYSMITH, B.J.; FURKERT, D.P.; BRIMBLE, M.A. Enduracididine, a rare amino acid component of peptide antibiotics: Natural products and synthesis. **Beilstein Journal of Organic Chemistry**, Frankfurt, v. 12, p. 2325–2342, 2016.

BACKER, L.C.; LANDSBERG, J.H.; MILLER, M.; KEEL, K.; TAYLOR, T.K. Canine cyanotoxin poisonings in the United States (1920s–2012): Review of suspected and confirmed cases from three data sources. **Toxins**, Basel, v. 5, p. 597-1628, 2013.

BECKER, V.; IHARA, P.; YUNES, J.S.; HUSZAR, V.L.M. Occurrence of anatoxin-a(S) during a bloom of *Anabaena crassa* in a water-supply reservoir in southern Brazil. **Journal of Applied Phycology**, Dordrecht, v. 22, p. 235–241, 2010.

BULLERJAHN, G.S.; MCKAY, R.M.; DAVIS, T.W.; BAKER, D.B.; BOYER, G.L.; D'ANGLADA, L.V.; DOUCETTE, G.J.; HO, J.C.; IRWIN, E.G.; KLING, C.L.; KUDELA, R.M.; KURMAYER, R.; MICHALAK, A.M.; ORTIZ, J.D.; OTTEN, T.G.; PAERL, H.W.; QIN, B.; SOHNGEN, B.L.; STUMPF, R.P.; VISSER. P.M.; WILHELM, S.W. Global solutions to regional problems: Collecting global expertise to address the problem of harmful cyanobacterial blooms. A Lake Erie case study. **Harmful Algae**, Amsterdam, v. 54, p. 223-238 2016.

BURFORD, M.A; CAREY, C.C.; HAMILTON, D.P.; HUISMAN, J.; PAERL, H.W.; WOOD, S.A.; WULFF, A. Perspective: Advancing the research agenda for improving understanding of cyanobacteria in a future of global change. **Harmful Algae**, Amsterdam, v. 91, 101601, 2020. doi: 10.1016/j.hal.2019.04.004.

BURROUGHS, A.M.; HOPPE, R.W.; GOEBEL, N.C.; SAYYED, B.H.; VOEGTLINE, T.J.; SCHWABACHER, A.W.; ZABRISKIE, T.M.; SILVAGGI, N.R. Structural and functional characterization of MppR, an enduracididine biosynthetic enzyme from *Streptomyces hygroscopicus*: functional diversity in the acetoacetate decarboxylase-like superfamily. **Biochemistry**, Washington, DC, v. 52, p. 4492-4506, 2013.

CARMICHAEL, W.W.; GORHAM, P.R. Anatoxins from clones of Anabaena flos-aquae isolated from lakes of Western Canada. **Mitteilungen Internationale Vereinigung für Theoretische und Angewandte Limnologie**, Heidelberg, v. 21, p. 285-295, 1978.

CARMICHAEL, W.W.; MAHMOOD, A.; HYDE, E.G. Natural toxins from cyanobacteria (blue-green) algae. In: HALL, S.; STRICHARTZ, G. (Ed.). **Marine toxins**: Origin, structure, and molecular pharmacology. Washington, DC: American Chemical Society, 1990. p. 87-106.

CHANG, C.Y.; LYU, S.Y; LIU, Y.C; HSU, N.S; WU, C.C.; TANG, C.F.; LIN, K.H.; HO, J.Y.; WU, C.J.; TSAI, M.D.; LI, T.L. Biosynthesis of streptolidine involved two unexpected intermediates produced by a dihydroxylase and a cyclase through unusual mechanisms. **Angewandte Chemie International Edition**, New York, v. 53, p. 1943-1948, 2014.

CHATZIEFTHIMIOU, A.D.; RICHER, R.; ROWLES, H.; POWELL, J.T.; METCALF, J. Cyanotoxins as a potential cause of dog poisonings in desert environments. **The Veterinary Record**, London, v. 174, p. 484–485, 2014.

CHORUS, I. (Ed.). **Current approaches to cyanotoxin risk assessment**, **risk management and regulations in different countries**. Berlin: Umweltbundesamt, 2012.

COOK, W.O.; BEASLEY, V.R.; DAHLEM, A.M.; DELLINGER, J.A.; HARLIN, K.S.; CARMICHAEL, W.W. Comparison of effects of anatoxin-a(S) and paraoxon, physostigmine and pyridostigmine on mouse brain cholinesterase activity. **Toxicon**, Oxford, v. 26, n. 8, p. 750–753, 1988.

COOK, W.O.; DAHLEM, A.M.; LOVELL, R.A.; DAHLEM, A.M.; HOOSER, S.B.; MAHMOOD, N.A.; CARMICHAEL, W.W. Consistent inhibition of peripheral cholinesterases by neurotoxins from the TT freshwater cyanobacterium *Anabaena flos aquae*: studies of ducks, swine, mice and a steer. **Environmental Toxicology and Chemistry**, New York, v. 8, p. 915-922, 1989a.

COOK, W.O.; DELLINGER, J.A.; SINGH, S.S.; DAHLEM, A.M.; CARMICHAEL, W.W.; BEASLEY, V.R. Regional brain cholinesterase activity in rats injected intraperitoneally with anatoxin-a(S) or paraoxon. **Toxicology Letters**, Amsterdam, v. 49, p. 29–34, 1989b.

FIORE, M.F.; LIMA, S.T.; CARMICHAEL, W.W.; MCKINNIE, S.M.K.; CHEKAN, J.R.; MOORE, B.S. Guanitoxin, re-naming a cyanobacterial organophosphate toxin. **Harmful Algae**, Amsterdam, v. 92, e101737, 2020. doi: 10.1016/j.hal.2019.101737.

FOSS, A.J.; AUBEL, M.T.; GALLAGHER, B.; METTEE, N.; MILLER, A.; FOGELSON, S.B. Diagnosing microcystin intoxication of canines: clinicopathological indications, pathological characteristics, and analytical detection in postmortem and antemortem samples. **Toxins**, Basel, v. 11, E456, 2019. doi: 10.3390/toxins11080456.

GORHAM, R.P. Laboratory studies on the toxins produced by waterblooms of blue-green algae. **American Journal of Public Health and the Nation's Health**, New York, v. 52, p. 2100-21005, 1962.

GUGGER, M.; LENOIR, S.; BERGER, C.; LEDREUX, A.; DRUART, J.; HUMBERT, J.F.; GUETTE, C.; BERNARD, C. First report in a river in France of the benthic cyanobacterium *Phormidium favosum* producing anatoxin-a associated with dog neurotoxicosis. **Toxicon**, Oxford, v. 45, p. 919–928, 2005.

HALTLI, B.; TAN, Y.; MAGARVEY, N.A.; WAGENAAR, M.; YIN, X.; GREENSTEIN, M.; HUCUL, J.A.; ZABRISKIE, T.M. Investigating β-hydroxyenduracididine formation in the biosynthesis of the mannopeptimycins. **Chemistry & Biology**, New York, v. 12, p. 1163-1168, 2005.

HAN, L.; SCHWABACHER, A.W.; MORAN, G.R.; SIVAGGI, N.R. *Streptomyces wadayamensis* MppP is a pyridoxal 5′-phosphate-dependent L-arginine α-deaminase, γ-hydroxylase in the enduracididine biosynthetic pathway. **Biochemistry**, Washington, DC, v. 54, p. 7029-7040, 2015.

HAN, L.; VUKSANOVIC, N.; OEHM, S.A.; FENSKE, T.G.; SCHWABACHER, A.W.; SILVAGGI, N.R. *Streptomyces wadayamensis* mppP a PLP-dependent oxidase, not an oxygenase. **Biochemistry**, Washington, DC, v. 57, n. 23, p. 3252-3264, 2018.

HARKE, M.J.; DAVIS, T.W.; WATSON, S.B.; GOBLER, C.J. Nutrient-Controlled Niche Differentiation of Western Lake Erie Cyanobacterial Populations Revealed via Metatranscriptomic Surveys. **Environmental Science & Technology**, Washington, DC, v. 50, n. 2, p. 604-615, 2016.

HELMETAG, V.; SAMEL, S.A.; THOMAS, M.G.; MARAHIEL, M.A.; ESSEN, L. Structural basis for the erythro-stereospecificity of the L-arginine oxygenase VioC in viomycin biosynthesis. **FEBS Journal**, Oxford, v. 276, p. 3669-3682, 2009.

HEMSCHEIDT, T.; BURGOYNE, D.L.; MOORE, R.E. Biosynthesis of anatoxin-a(S). (2S,4S)-4-hydroxyarginine as an intermediate. **Journal of the Chemical Society, Chemical Communications**. London, n. 2, p. 205-206, 1995.

HENRIKSEN, P.; CARMICHAEL, W.W.; NA, J.; MOESTRUP, O. Detection of an anatoxin-a(S)-like anticholinesterase in natural blooms and cultures of cyanobacteria/bluegreen algae from Danish lakes and in the stomach contents of poisoned birds. **Toxicon**, Oxford, v. 35, n. 6, p. 901-913, 1997.

HYDE, E.G.; CARMICHAEL, W.W. Anatoxin-a(S), a naturally occurring organophosphate, is an irreversible active site-directed inhibitor of acetylcholinesterase (EC3.1.1.7). **Journal of Biochemical Toxicology**, New York, v. 6, p. 195-201, 1991.

JOCHIMSEN, E.M.; CARMICHAEL, W.W.; AN, J.; CARDO, D.M.; COOKSON, S.T.; HOLMES, C.E.M.; ANTUNES, M.B.; DE MELO FILHO, D.A.; LYRA, T.M.; BARRETO, V.S.T.; AZEVEDO, S.M.F.O.; JARVIS, W.R. Liver failure and death after exposure to microcystins at a hemodialysis center in Brazil. **The New England Journal of Medicine**, Boston, v. 338, p. 873-878, 1998.

KUMAGAI, T.; OZAWA, T.; TANIMOTO, M.; NODA, M.; MATOBA, Y.; SUGIYAMA, M. High-level heterologous production of D-cycloserine by *Escherichia coli*. **Environmental Microbiology**, Oxford, v. 81, p. 7881–7887, 2015.

KUMAGAI, T.; TAKAGI, K.; KOYAMA, Y.; MATOBA, Y.; ODA, K.; NODA, M.; SUGIYAMA, M. Heme protein and hydroxyarginase necessary for biosynthesis of D-cycloserine. **Antimicrobial Agents and Chemotherapy**, Washington, DC, v. 56, p. 3682-3689, 2012.

MAHMOOD, N.A.; CARMICHAEL, W.W. The pharmacology of anatoxin-a(S), a neurotoxin produced by the fresh-water cyanobacterium *Anabaena flos-aquae* NRC 525-17. **Toxicon**, Oxford, v. 24, p. 425-434, 1986.

MAHMOOD, N.A.; CARMICHAEL, W.W. Anatoxin-a(S), an anticholinesterase from the cyanobacterium *Anabaena flos-aquae* NRC-525–17. **Toxicon**, Oxford, v. 25, n. 11, p. 1221-1227, 1987.

MAHMOOD, N.A.; CARMICHAEL, W.W.; PFAHLER, D. Anticholinesterase poisoning in dogs from cyanobacterial (blue green algae) bloom dominated by *Anabaena flos-aquae*. **American Journal of Veterinary Research**, Chicago, v. 49, n. 4, p. 500–503, 1998.

MATSUNAGA, S.; MOORE, R.E.; NIEMCZURA, W.P.; CARMICHAEL, W.W. Anatoxina(s), a potent anticholinesterase from *Anabaena flos-aquae*. **Journal of the American Chemical Society**, Washington, DC, v. 111, n. 20, p. 8021–8023, 1989.

METCALF, J.S.; MORRISON, L.F.; KRIENITZ, L.; BALLOT, A.; KRAUSE, E.; KOTUT, K.; PÜTZ, C.; PFLUGMACHER, W.S.; CODD, G.A. Analysis of the cyanotoxins anatoxin-a and microcystins in Lesser Flamingo feathers. **Toxicological and Environmental Chemistry**, New York, v. 88, p. 159–167, 2006.

MOLICA, R.J.R.; OLIVEIRA, E.J.A.; CARVALHO, P.V.V.C.; COSTA, A.N.S.F.; CUNHA, M.C.C.; MELO, G.L.; AZEVEDO, S.M.F.O. Occurrence of saxitoxins and an anatoxin-a(S)-like anticholinesterase in a Brazilian drinking water supply. **Harmful Algae**, Amsterdam, v. 4, n. 4, p. 743–753, 2005.

MONSERRAT, J.M.; YUNES, J.S.; BIANCHINI, A. Effects of *Anabaena spiroides* (Cyanobacteria) aqueos extracts on the aceltycholinesterase activity of aquatic species. **Environmental Toxicology and Chemistry**, New York, v. 20, p. 1228-1235, 2001.

MOORE, B.S.; OHTANI, I.; DE KONING, C.B.; MOORE, R.E.; CARMICHAEL, W.W. Biosynthesis of anatoxin-a(S). Origin of the carbons. **Tetrahedron Letters**, Oxford, v. 33, p. 6595-6598, 1992.

MOORE, R.E.; OHTANI, I.; MOORE, B.S. DE KONING. C.B.; YOSHIDA, W.Y.; RUNNEGAR, M.T.C.; CARMICHAEL, W.W. Cyanobacterial toxins. **Cazzetta Chimica Italiana**, Rome, v. 123, p. 329-336, 1993.

NEGRI, A.P.; JONES, G.J.; HINDMARSH, M. Sheep mortality associated with paralytic shellfish poisons from the cyanobacterium *Anabaena circinalis*. **Toxicon**, Oxford, v. 33, p. 1321–1329, 1995.

ONODERA, H.; OSHIMA, Y.; HENRIKSEN, P.; YASUMOTO, T. Confirmation of anatoxin-a(S), in the cyanobacterium *Anabaena lemmermannii*, as the cause of bird kills in Danish lakes. **Toxicon**, Oxford, v. 35, n. 11, p. 1645–1648, 1997.

OTTEN, T.G.; PAERL, H.W. Health effects of toxic cyanobacteria in U.S. drinking and recreational waters: our current understanding and proposed direction. **Current Environmental Health Reports**, Geneva, v. 2, p. 75–84, 2015.

PEARSON, L.A.; DITTMANN, E.; MAZMOUZ, R.; ONGLEY, S.E.; D'AGOSTINO, P.M.; NEILAN, B.A. The genetics, biosynthesis and regulation of toxic specialized metabolites of cyanobacteria. **Harmful Algae**, Amsterdam, v. 54, p. 98-111, 2016.

THOMAS, M.G.; CHAN, Y.A.; OZANICK, S.G. Deciphering tuberactinomycin biosynthesis: isolation, sequencing, and annotation of the viomycin biosynthetic gene cluster. **Antimicrobial Agents and Chemotherapy**, Washington, DC, v. 47, n. 9, p. 2823–2830, 2003.

WERNER, V.R.; LAUGHINGHOUSE IV, H.D.; FIORE, M.F.; SANT'ANNA, C.L.; HOFF, C., SANTOS, K.R.D.S.; NEUHAUS, E.B.; MOLICA, R.J.R.; HONDA, R.Y.; ECHENIQUE, R.O. Morphological and molecular studies of *Sphaerospermopsis torques-reginae* (Cyanobacteria, Nostocales) from South American water blooms. **Phycologia**, Berkeley, v. 51, p. 228–238, 2012.

4 CONCLUSION

The results presented in this thesis achieved the goals proposed for the project. Our work, and the way it was conducted during the doctoral period, was able to confirm the initial hypothesis for this thrilling project: the validation of the biosynthetic gene cluster proposed to produce the neurotoxin guanitoxin.

Using heterologous expression, purification and characterization of proteins we were able to test and demonstrate enzymatic activities of the nine biosynthetic GNT enzymes and assemble the complete biosynthetic pathway. The international internship in the laboratory of Prof. Dr. Bradley S. Moore was crucial for this achievement due to the expertise of Dr. Moore and his research team in enzymology and natural products biosynthetic pathways.

This doctoral thesis shows the elucidation of a very potent neurotoxin produced by a bloom-forming cyanobacteria, which is been overlooked by the scientific community and water managers authorities. Thus, we expect that this project may generate a patent for development of sensitive and rapid methods of environmental detection and, also, call the attention to this powerful and water-soluble neurotoxin which could be present in freshwaters used for public supply.

APPENDICES

Appendix A: Supplementary Material of Chapter III

Supplementary Materials for

Disentangling the biosynthesis of a neurotoxic natural organophosphate in Cyanobacteria

Stella T. Lima, Jonathan R. Chekan, Shaun M. K. McKinnie, Danillo O. Alvarenga, Hanna Luhavaya, Felipe A. Dörr, Augusto Etchegaray, Ernani Pinto, Hans Renata, Bradley S. Moore*, Marli F. Fiore*

Correspondence to: fiore@cena.usp.br (M.F.F); bsmoore@ucsd.edu (B.S.M).

Supplementary Materials includes:

General Material and Methods

Supplementary Figures: Figs. S1 to S13

Chemical Synthesis

Supplementary Table: Table S1

General materials and methods

Chemically Competent *E. coli* cells

The competent cells (DH10β and BL21(DE3)) were produced using an overnight *E. coli* culture as an inoculum (1% v/v) into 50 mL of LB medium. The culture was incubated at 37 °C, 220 rpm until the OD_{600} was 0.3-0.5. The cells were harvested by centrifugation at 2,200 x g, 5 minutes at 4 °C, and suspended in 30 mL of ice-cold Buffer A (CH_3COOK 10mM, $CaCl_2$ 50mM, pH 6.2). After 1 hour of incubation on ice with occasional agitation, the cells were spun down at 2,200 x g for 5 minutes at 4 °C and suspended in 3 mL ice-cold Buffer B (CH_3COOK 10mM, $CaCl_2$ 50mM, Glycerol 20% (v/v), pH 6.2). Aliquots of 100 µL were transferred into pre-cooled tubes, flash-frozen in liquid nitrogen and stored at –80 °C.

Transformation into *E. coli*

The plasmids were transformed into *E. coli* DH10β chemically competent cells for storage and BL21(DE3) for expression. The transformation by heat shock proceeded according to the following protocol: 0.5 µL of plasmid was add to the chemical competent cells and maintained into ice for 30 minutes. After this, the cells were heated to 42 °C for 45 seconds and placed in the ice again for 3 minutes; 900 µL of LB medium was added in the tube and the cells were incubated for 50 minutes at 37 °C and 200 rpm of agitation. After this step, the cells were

plated on LB agar plates supplemented with the corresponding antibiotics. The plates were incubated at 37 °C, overnight. An inoculum with colonies was grown overnight in the same condition of LB agar plates at 37 °C and 200 rpm of agitation and purified following the protocol of Plasmid DNA Purification QIAprep Spin Miniprep Kit (QIAGEN). After the purification the plasmid, concentrations were measured by NanoDrop® and they were stored at –20 °C.

Extraction and LC-MS analyses

Lyophilized culture was extracted with 5 mL of ethanol/acetic acid 0.1M (20:80 v/v), sonicated for 1 minute on ice and centrifuged at 5,000 xg for 15 minutes. The supernatant was lyophilized and resuspended in methanol and filtered into an autosampler vial. The Hydrophilic Interaction Liquid Chromatography (HILIC) separation was carried out on a Zic-Hilic, 150x2 mm, 5 µm, 200 Å (Merck) column.

Method A: Separation was achieved under gradient elution at 0.2 mL/min where elution A was 5 mM ammonium formate containing 0.01% of formic acid, and elution B was acetonitrile/water (90:10 v/v) with 0.01% of formic acid as well. Elution started with a linear gradient of 90% B to 20% until 35 min, second isocratic gradient of 20% B until 37.50 min and a third isocratic gradient of 90% B until 45 min.

Method B: For this analysis a Synergi Polar-RP 4µ 250 x 4.6 mm column used at 0.75 mL/min with the following method: 0% B (12 min), 0 to 100% B (5 min), 100% B (3 min), 100 to 0% B (2.5 min), 0% B (2.5 min), wherein A = 0.1% aqueous formic acid, and B = 0.1% formic acid in acetonitrile.

DNA extraction

S. torques-reginae ITEP-024 cells were washed according the protocol: 50 mL oh MilliQ water twice, followed by Extran 0.05%, wash solution twice (NaCl 50 mM, Tris-HCl 10 mM pH 7.5, EDTA 2.5 mM pH 8.0) and, lastly, NaCl 0.85%. Every step was harvested by centrifugation at 5,000 xg for 10 minutes at room temperature. Washed cells were placed on mortar and poured liquid N_2 into it. Gently, the cells were crushed to powder with pestle as liquid N_2 evaporated. Cyanobacteria cells were scraped into a 2 mL microtube with plastic scoop and DNA extraction was proceeded according QIAGEN Bacteria Genomic DNA extraction protocol (QIAGEN, Germany).

PCR methods

Coexpression of the GntB and GntC into pCOLADuet-1 vector was carried out amplifying codon optimized gene for expression in *E. coli* synthesized and commercialized by GenScript Inc. (Piscataway, NJ, USA). In one of the multiple cloning site (MCS) was inserted *gnt*B (957 bp) and in the other MSC *gnt*C (1113 bp). The plasmid was also amplified for Gibson Assembly Master Mix (New England Biolabs *Inc*.) to prevent the plasmid from assembling without the fragments of interest. For primer set gntB-F/R program used was 98 °C for 30 seconds for denaturing conditions, 30 cycles were completed with 98 °C for 10 seconds, 62 °C for 30 seconds and 72 °C for 30 seconds, followed by 72 °C for 2 minutes for final extension. For primer set gntC-F/R program used was 98 °C for 30 seconds for denaturing conditions, 30 cycles of 98 °C for 10 seconds, 70 °C for 30 seconds and 72 °C for 30 seconds and additionally 72 °C for 3 minutes for final extension. For primer sets pCOLADuet-F/R the program used was 98 °C for 30 seconds for denaturing conditions, 30 cycles of 98 °C for 10 seconds, 61 °C constantly for 30 seconds and 72 °C for 2 minutes, and finally 72 °C for 2 minutes for final elongation.

Vector assembly and transformation

For the vector assembly, the fragments were amplified following the conditions and the programs specified. All the fragments were loaded in a 0.7% agarose gel with Gel Loading Dye Purple (6X) (New England Biolabs Inc.) and run at 90 volts for 50 minutes. The bands were cut, and the DNA was extract with QIAquick Gel Extraction Kit Protocol (QIAGEN). The DNA Clean & Concentrator Kit (Zymo Research) was used to clean and concentrate the fragments, reaching high enough concentration for the assembly. The Gibson Assembly Master Mix (New England Biolabs *Inc*.) was used to assemble Duet plasmid, resulting in two DNA fragments into a unique vector. The reaction was incubated at 50 °C for 1 hour. After that, the assembled vectors were transformed into chemical competent cells. For Duet vectors, DH10β was used to transform and replicate the vector. 5 µL of Duet Gibson Assembly product was added to 100 µL of cells and 5 µL of empty plasmid was also added as a positive control following the transformation protocol described before. Plasmids were confirmed by Sanger Sequencing by Integrated DNA Technologies, Inc. (San Diego, CA, USA).

Vector Expression

For expression, the selected plasmid was transformed into *E. coli* BL21(DE3) following the same protocol for chemical transformation described before. The expression assay was

carried out at 18 °C in M9 Minimal media, 30 µg/mL of kanamycin, 200 rpm of agitation. The flasks were placed at 37 °C until the OD_{600} of 0.7 at which point flasks were moved to 18 °C. After 1 hour, 1 mM of IPTG was added and the expression flasks were incubated for 5 days under constantly agitation. Cells were harvested by centrifugated and the clarified media was collected.

Enzymes heterologous expression

The enzymes coding sequences were optimized for expression in *E. coli* and synthesized by GenScript Inc. (Piscataway, NJ, USA). The synthetic genes of guanitoxin (GNT) pathway were sub cloned into the pET28a(+) kanamycin resistant expression vector and contained a N-terminal histidine tag (GenScript Inc.). The pET28a(+) vectors containing the synthetic genes were resuspended in 100 µL of sterilized ultrapure water and 0.5 µL of the plasmids were transformed into 100 µL of chemically competent *E. coli* DH5α cells for plasmid replication and storage, and 0.5 µL were transformed into 100 µL of BL21(DE3) cells for the expression experiments, as described before. Pre inoculums for expression were prepared by taking colonies from the Petri dishes and inoculating it into tubes containing 10 mL of LB medium with 50 µg/mL of kanamycin. The tubes were incubated at 37 °C under agitation of 250 rpm, overnight. Each pre inoculum was poured into 2 L flasks containing 1 L of TB media supplemented with the same concentration of antibiotic, and incubated at 37°C, under agitation of 200 rpm. When the cultures reached an OD_{600} of ~0.9, the temperature was dropped to 18°C and after 1 hour the protein expression was induced by adding 1 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG). The cultures were incubated at 18 °C, overnight, with shaking at 200 rpm. The expression culture was transferred into appropriate tubes and centrifuged for 8000 x g at 4 °C for 10 min. The cell pellets were resuspended in Suspension Buffer (500 mM NaCl, 10% glycerol, 20 mM Tris pH 8.0) and stored at -80°C until purification.

Protein purification

The homogenized cells were transferred to a beaker and cells were lysed by sonication for a total of 6 min with pulses of 15 sec on, 45 sec off at 40% of amplitude. The temperature was maintained at 4 °C by keeping the beaker in an ice bath. The lysate was clarified by centrifugation at 15,000 x g for 30 min at 4°C. For GntB, a membrane protein, cells lysate and centrifugation steps were preceded following de Rond et al., 2017. GntB was sonicate in cycles of 1 min on, 1 min off at 25% of amplitude for 30 minutes. The lysate was then clarified by 8,000 x g for 10 min at 4°C (de Rond et al., 2017). The soluble proteins were purified initially

through affinity chromatography with a HisTrap FF (5 mL) column (GE Healthcare Life Sciences). The column was equilibrated with Wash Buffer (1M NaCl, 30 mM imidazole, 20 mM Tris pH 8.0). Around 30 mL of protein sample were applied at a flow rate of 2 mL min^{-1} to isolate the His_6-Gnt proteins. The soluble fraction of the lysate was applied in the column using ÄKTA purifier (GE Healthcare Life Sciences). After the application, the column was washed with 40 mL of Wash Buffer. The bound protein in the column was eluted using a linear gradient to 100% Elution Buffer (1 M NaCl, 250 mM imidazole, 20 mM Tris pH 8.0) over 40 mL while collecting 5 mL fractions. For protein GntC the method was the same, however instead of Wash Buffer and Elution Buffer, Buffer A (25 mM Tris pH 8.0, 300 mM NaCl, 100 mM imidazole, 100 μM PLP) and Buffer B (25 mM Tris pH 8.0, 300 mM NaCl, 250 mM imidazole) were used. Fractions were assessed for purity using SDS-PAGE, and fractions that were at least 90% pure were combined. The proteins fractions were collected and concentrated using Amicon Ultra-15 (Millipore), 30K or 10K, based in the size of protein, until approximately 2 mL. The protein was further purified by size exclusion chromatography using a HiLoad 16/60 Superdex 75 column or HiLoad 16/60 Superdex 200 column, based on proteins sizes and possibility of dimers (GE Healthcare Life Sciences). For this purification, the buffer used had 20 mM HEPES (pH 7.5) and 300 mM KCl, for GntA, D, E, F, G, I and J, and 100 mM MOPS (pH 6.7) and 20 μM PLP, for GntC. The fraction with desired protein was collected, pooled and concentrated in Amicon Ultra-15 (Millipore), 30K or 10K, based in the size of protein. Concentrations was determined by Bradford assay (Protein Assay Dye Reagent Concentrate, Bio-Rad). The purified proteins were stored in 50 μL aliquots, for crystal assays, and 20 μL aliquots, for enzymatic assays, at -80 °C.

Enzymatic assays

Following the proposed sequence for enzymatic activity against the proposed substrates, the reactions were proceeded as described below.

GntD: enduracididine β-hydroxylase

The GntD enzymatic reaction was carried out in the pCOLADuet_*gnt*BC clarified media in a total volume of 500 μL. For this activity enzymatic assay 20 μM GntD, 2 mM α-ketoglutarate, 0.1 mM $FeSO_4$ and 0.1 mM ascorbic acid was added. The reaction was incubated for 18 hours at room temperature. The reaction was quenched with one volume of acetonitrile (ACN) on ice and filtered at 14000 xg, 4 °C for 10 minutes using both 0.2 μM filters and 3kDa

cutoff filters, to remove cellular proteins and debris. Supernatant was then removed and subjected to LC-MS analysis.

GntG: aldolase

The GntG enzymatic reaction were set up using the filtrate of the pCOLADuet_*gnt*BC media + GntD assay. For this assay 20 μM GntG and 50 μM PLP was added. The reaction was incubated at 27 ºC for 18 hours, in a total volume of 300 μL, and quenched with one volume of acetonitrile (ACN) on ice and filtered at 14000 xg, 4 °C for 10 minutes using both 3kDa cutoff filters and 0.2 μM filters. Supernatant was then removed and subjected to LC-MS analysis.

GntE: aminotransferase

The GntE enzymatic reaction were set up in the filtrate of the pCOLADuet_*gnt*BC media + GntD+ GntG reaction. For this assay 20 μM GntE, 50 μM PLP and 10 mM glutamate was added. The reaction was incubated at 27 ºC for 18 hours, in a total volume of 100 μL, and quenched with one volume of acetonitrile (ACN) on ice and filtered at 14000 xg, 4 °C for 10 minutes using both 3kDa cutoff filters and 0.2 μM filters. Supernatant was then removed and subjected to LC-MS analysis.

GntF: N-methyltransferase

The GntF enzyme was tested with the synthesized primary amine substrate **(6)**. The reaction was completed with 20 μM GntF, 1 mM S-adenosyl methionine (SAM) in 50 mM of Tris pH 7.4 and incubated at 27 °C for 18 hours, in a total volume of 500 μL. The reaction was quenched with one volume of acetonitrile (ACN) on ice and filtered at 14000 xg, 4 °C for 10 minutes using both 3kDa cutoff filters and 0.2 μM filters. Supernatant was then removed and subjected to LC-MS analysis.

GntA: N-hydroxylase

The heme protein, GntA, was tested with the synthesized dimethylamine compound, intermediate (7). 20 μM enzyme was added to 5 mM of NADPH and 0.1 mM of (7), in 50 mM of Tris buffer pH 7.4. A total volume of 500 μL of reaction was incubated at 27 °C for 18 hours. The reaction was quenched with one volume of acetonitrile (ACN) on ice and filtered at 14000 xg, 4 °C for 10 minutes using 0.2 μM filters. Supernatant was then removed and subjected to LC-MS analysis.

GntI: kinase

The filtrate of GntA enzymatic assay was used to evaluate GntI's ability to phosphorylate the product of the GntA reaction, intermediate (8). 2 mM ATP, 100 mM NaCl, and 2 mM $MgCl_2$ was added to the filtrate, followed by 20 μM kinase GntI. The reaction volume was completed to 200 μL with 50 mM of Tris buffer pH 7.4 and incubated at 37 °C for 30 minutes. Part of the reaction mixture was quenched with one volume of acetonitrile (ACN) on ice and filtered at 14000 xg, 4 °C for 10 minutes using 0.2 μM filters. Supernatant was then removed and subjected to LC-MS analysis.

GntJ: O-methyltransferase

For the last proposed enzymatic reaction of GNT pathway, the production of guanitoxin; 1 mM of SAM was added to the filtrate of previous reaction, followed by 20 μM of GntJ enzyme without the His_6 tag. A total volume of 100 μL was completed with 50 mM of Tris buffer pH 7.4 and incubated at 27°C for 18 hours.

Searching for GNT biosynthetic gene cluster in metagenomics data base

The search for GNT biosynthetic genes was carried out using the metagenomic data base deposited in NCBI (National Center for Biotechnology Information). The guanitoxin genes sequences were used against the Metaomics data base from NCBI, using the tool BLAST (Basic Local Alignment Search Tool).

Supplementary figures

Fig. S1 - *S. torques-reginae* ITEP-024 bloom localization during end of summer and beginning of fall 2002, Tapacurá Reservoir, Recife, PE - Brazil (Molica et al., 2005). Map data generated using Map data © 2020 Google, INEGI (https://www.google.com/maps).

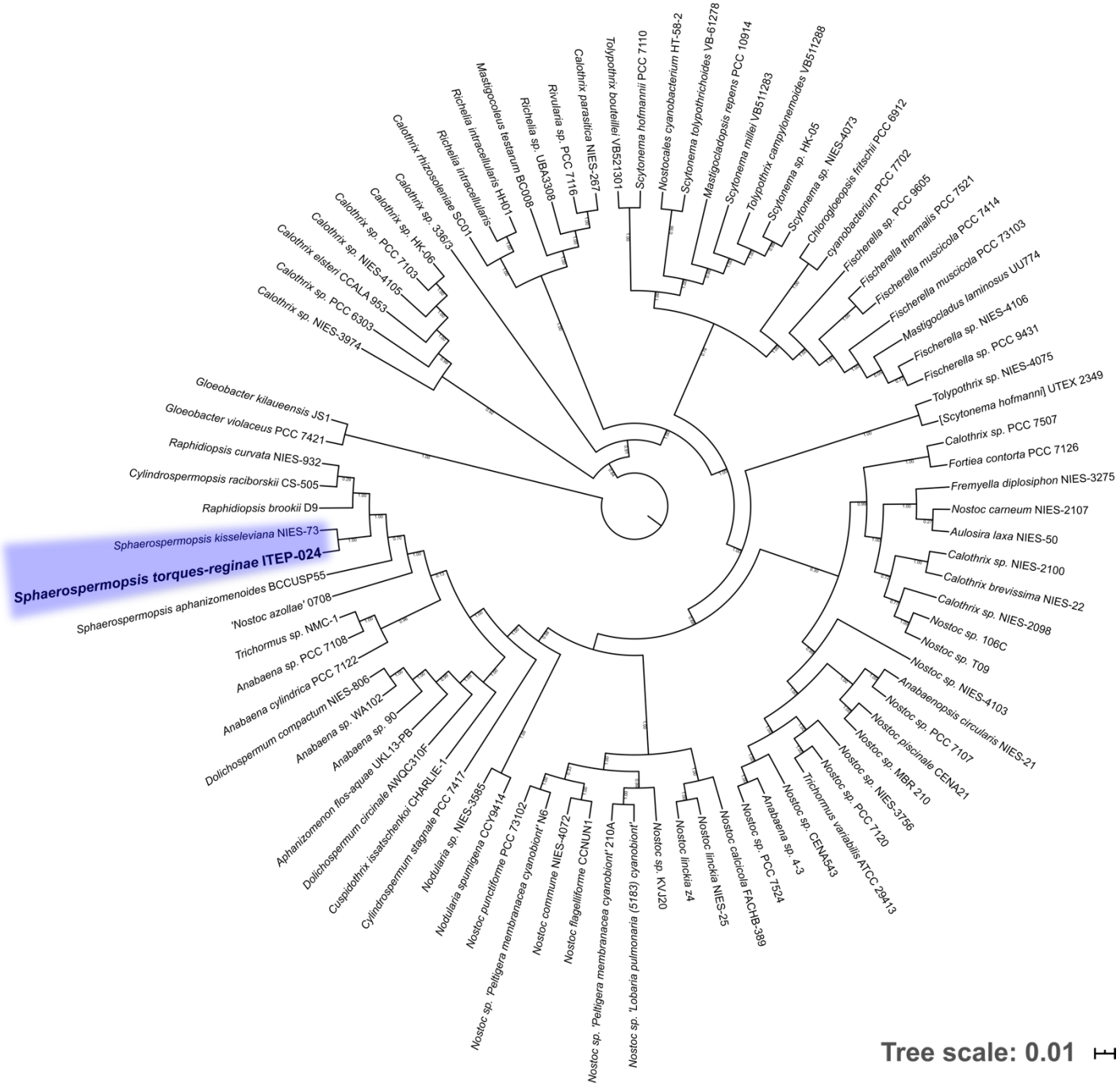


Fig. S2 - Genome phylogeny for *Sphaerospermopsis torques-reginae* ITEP-024. The genome tree is inferred using GTDB-Tk (PARKS et al., 2018) by approximately-maximum-likelihood phylogenetic analysis from an aligned concatenated set of 120 single copy marker proteins for Bacteria using Genome Taxonomy Database (GTDB).

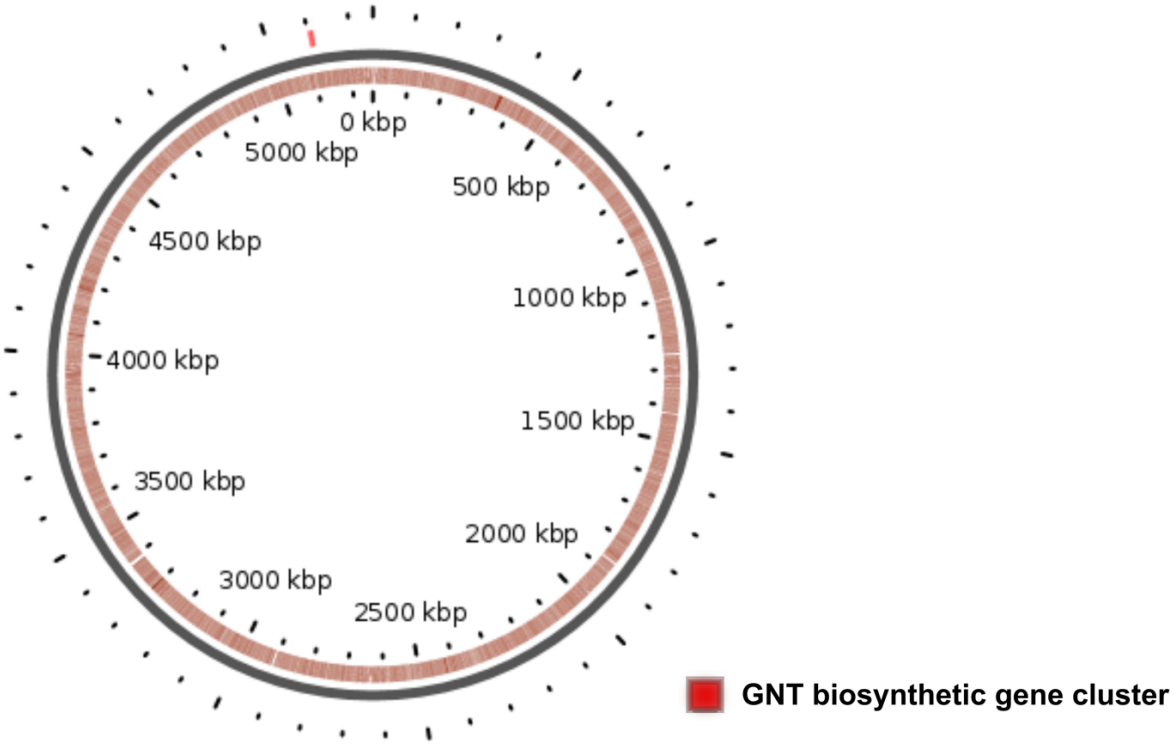


Fig.S3 - *Sphaerospermopsis torques-reginae* ITEP-024 5.2 kbp genome. Guanitoxin biosynthetic gene cluster is localized in the red bar.

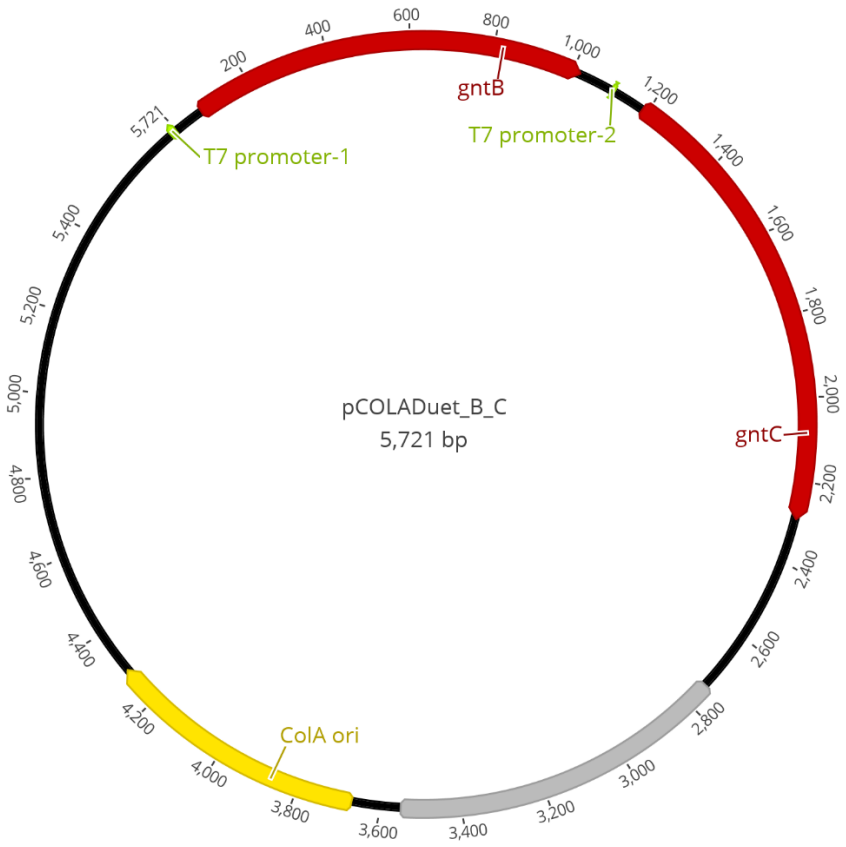


Fig. S4 - pCOLADuet-1 vector assembled with *gnt*B and *gnt*C genes (red) from the guanitoxin pathway for enduracididine production. The vector is designed for the coexpression of two target genes from a single plasmid, which encodes two multiple cloning sites (MCS) each of which is preceded by a T7 promoter (green), *lac* operator and ribosome binding site. The vector has the COLA replicon from ColA *ori* (yellow) and kanamycin resistance gene (gray).

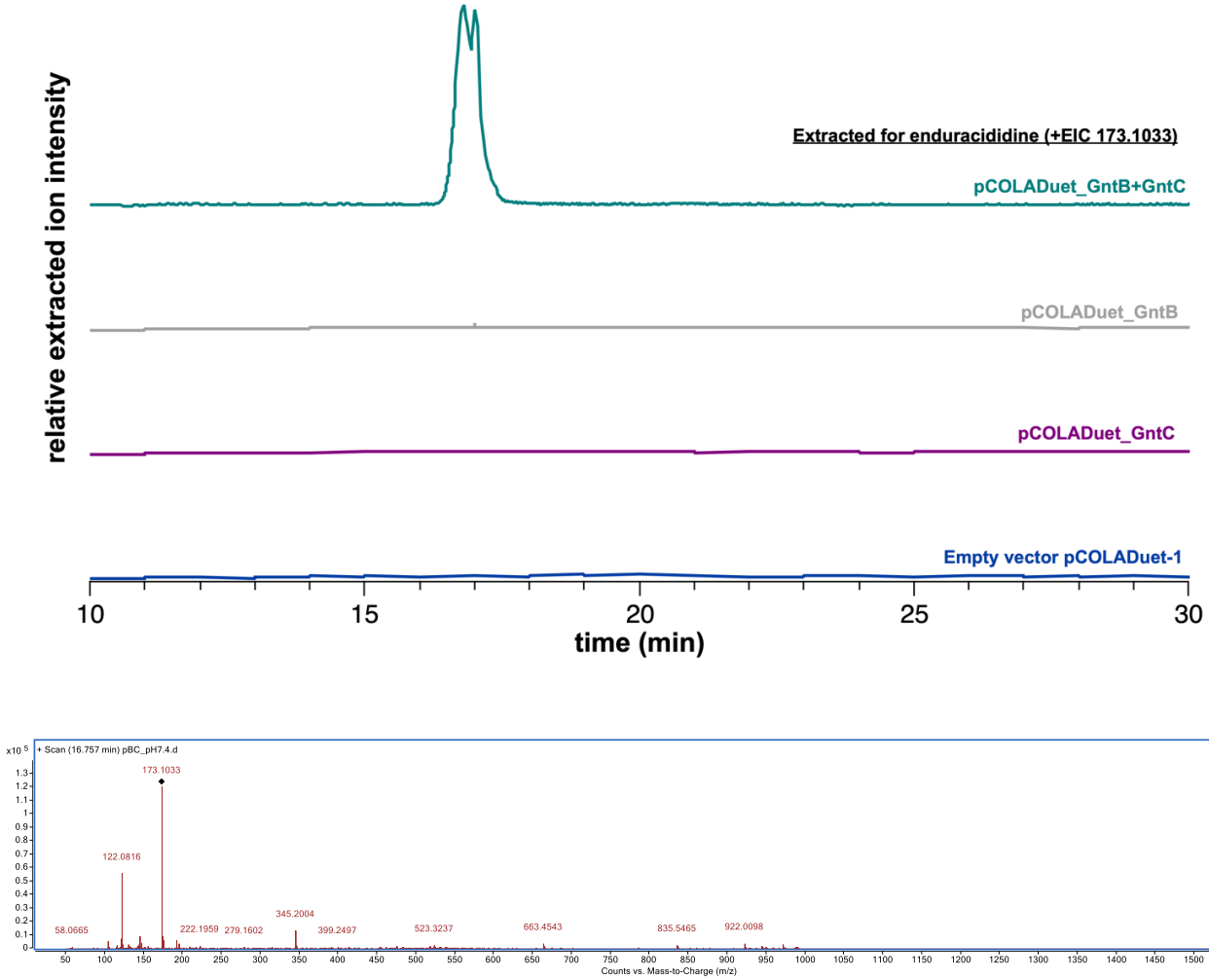


Fig. S5 - Positive mode LCMS chromatograms of pCOLADuet_GntB_GntC *in vivo* using Method A, showing the extracted ion chromatograms (EIC ± 0.01 *m/z*) for enduracididine. The expression was set up as previously described. Positive mode LCMS chromatograms of pCOLADuet_GntB_GntC culture expression extract shows enduracididine peak (+EIC 173.1033 ± 0.01 m/z, cyan trace).

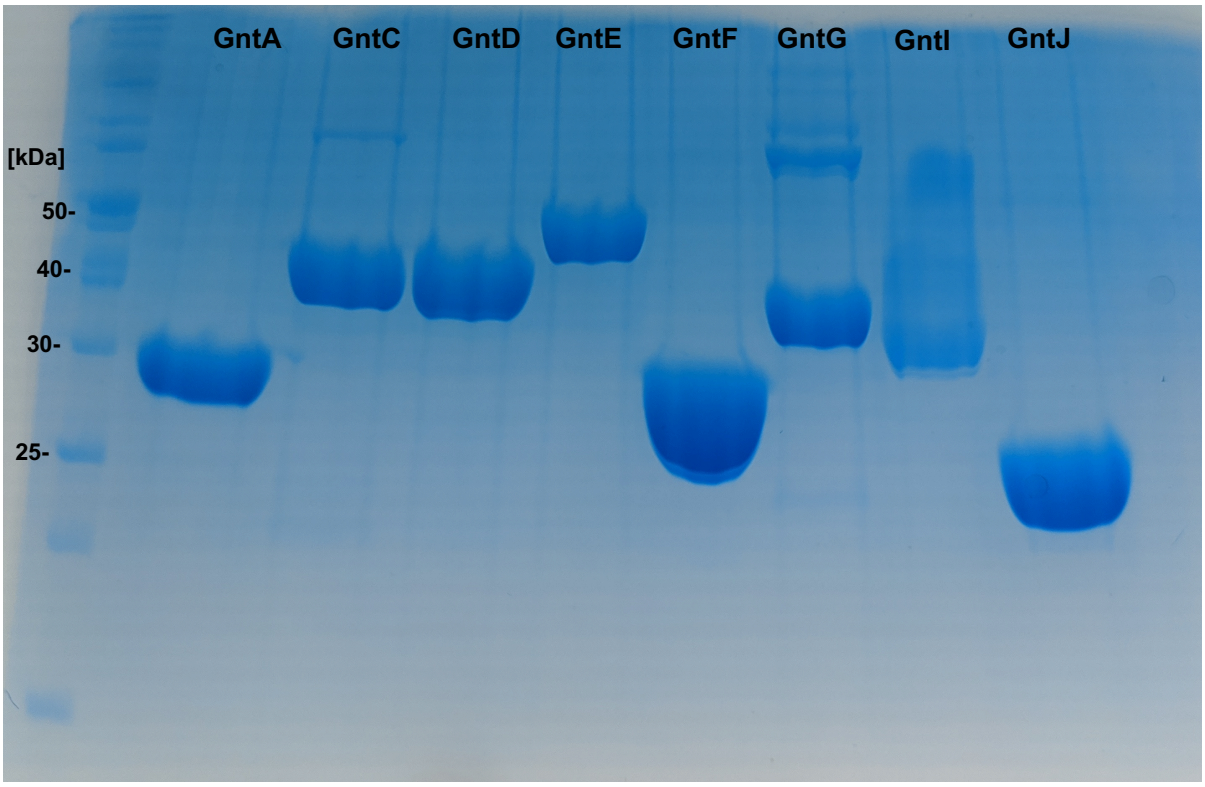
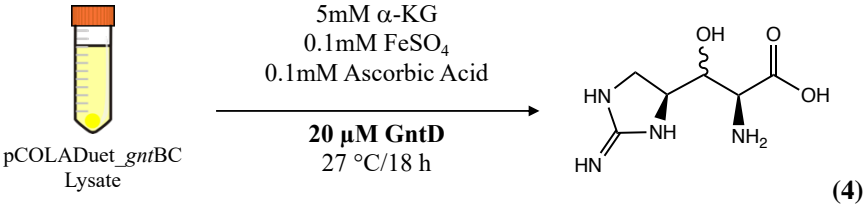
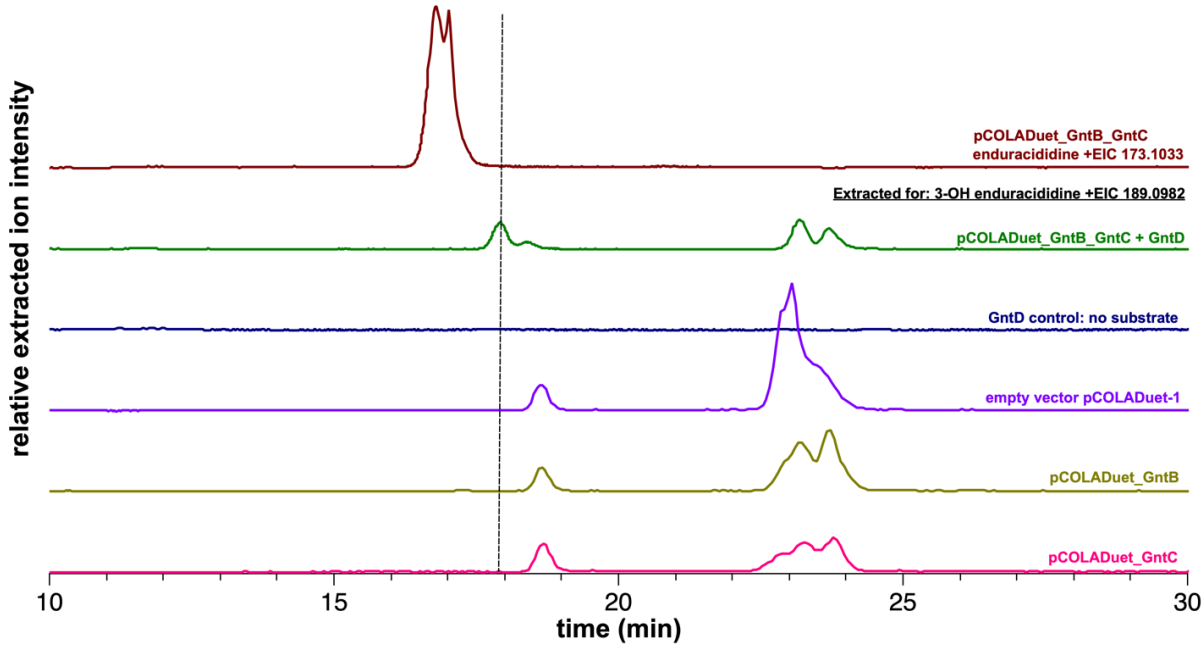


Fig. S6 - 12% SDS-polyacrylamide gel for purified GNT proteins with EZ-Run *Rec* Protein Ladder (Fisher Bioreagents).



(4)



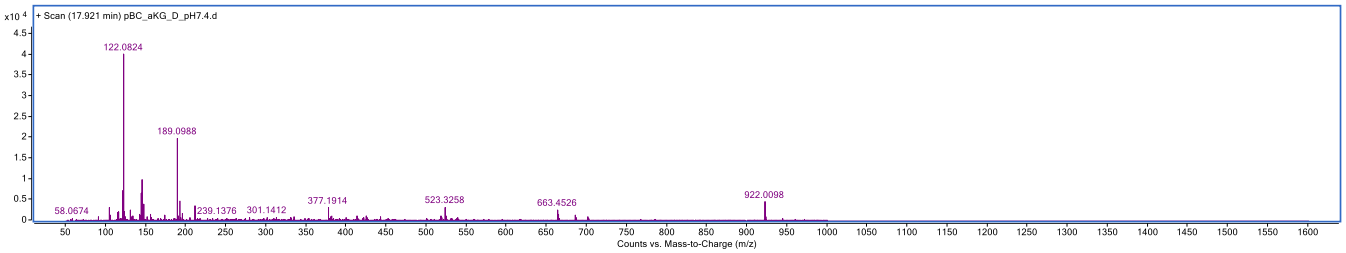


Fig. S7 - GntD reactions were set up as described before using pCOLADuet_GntB_GntC extracted and incubated for 18 hours at room temperature. Positive mode LCMS chromatograms using LC Method A, showing the extracted ion chromatogram for anticipated product (EIC ± 0.01 m/z). Positive mode LCMS chromatogram of GntD with pCOLADuet_GntB_GntC lysate extract shows β-OH enduracididine peak (+EIC 189.0988 ± 0.01 m/z, green trace).

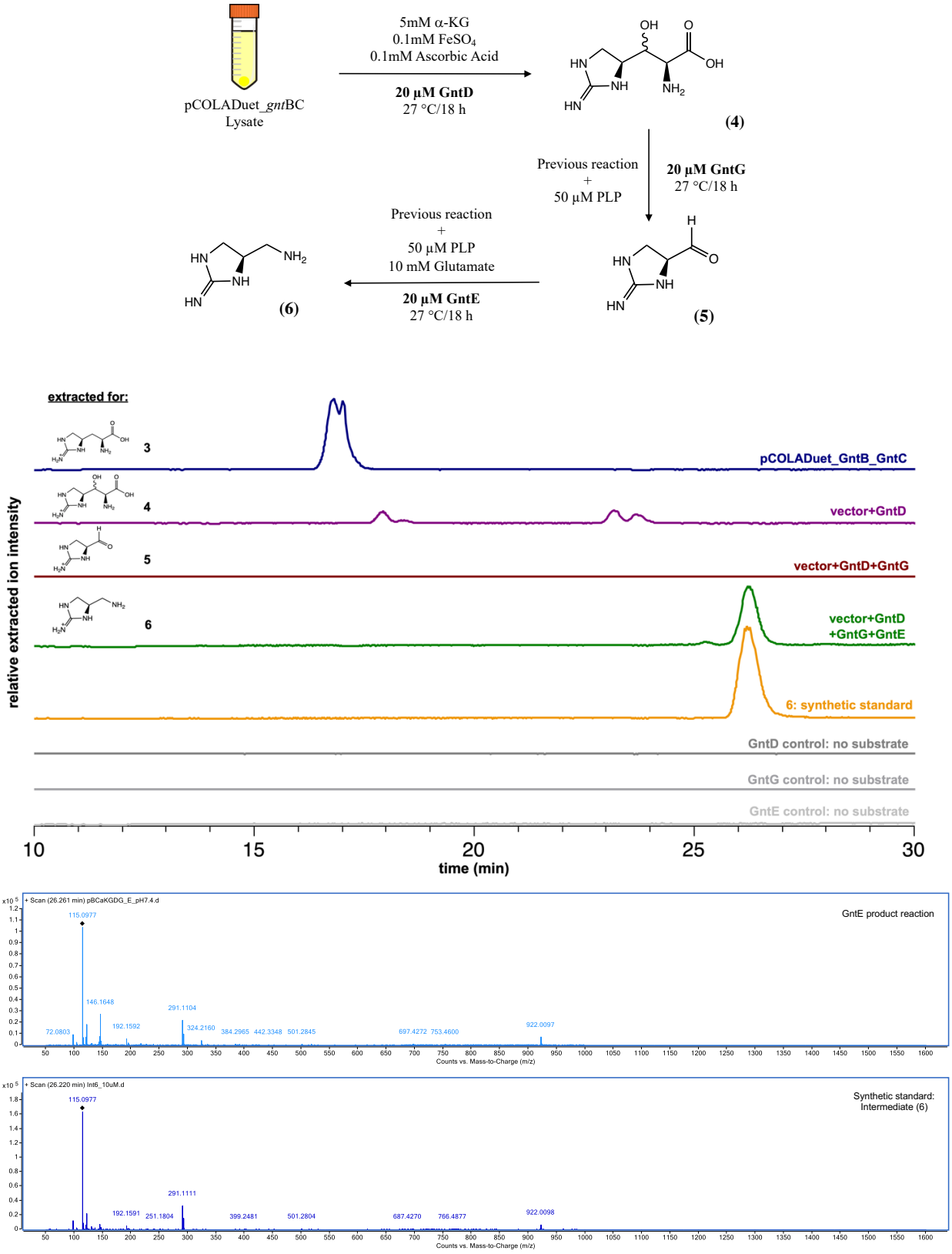


Fig. S8 - First half of GNT pathway combined reactions of *in vivo* pCOLADuet_GntB_GntC vector and *in vitro* purified enzymes covering steps from L-arginine to (6). Reactions were set up as previously described. Positive mode LC-MS chromatograms using LC Method A, showing the extracted ion chromatogram for anticipated product (EIC ± 0.01 m/z). The expected GntG reaction product, (5), was not found likely because it is unstable in solution. Positive mode LCMS chromatogram of GntE with previous reactions shows intermediate (6) peak (+EIC 115.0977 ± 0.01 *m/z*, green trace) and synthetic standard for same compound (+EIC 115.0977 ± 0.01 *m/z*, orange trace).

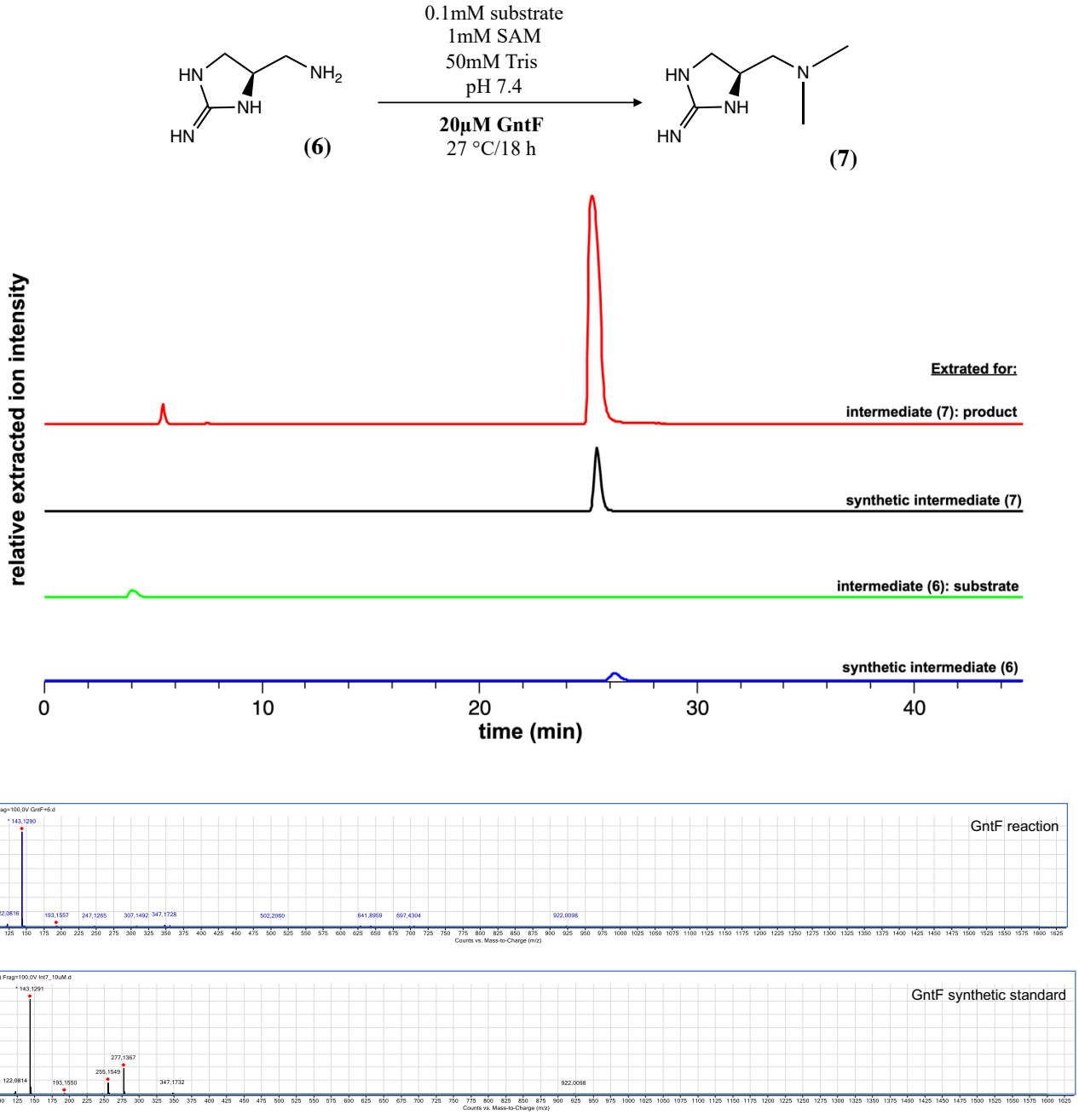


Fig. S9 - GntF reaction were set up as described before using 1 mM of SAM and incubated for 18 hours at 27 ºC. Positive mode LCMS chromatograms using LC Method A, showing the extracted ion chromatogram for product, synthetic product standard, substrate, and synthetic substrate standard (EIC ± 0.01 *m/z*). Positive mode LCMS chromatogram of GntF reaction product shows the peak for (7) (+EIC 143.0990 ± 0.01 *m/z*, red trace).

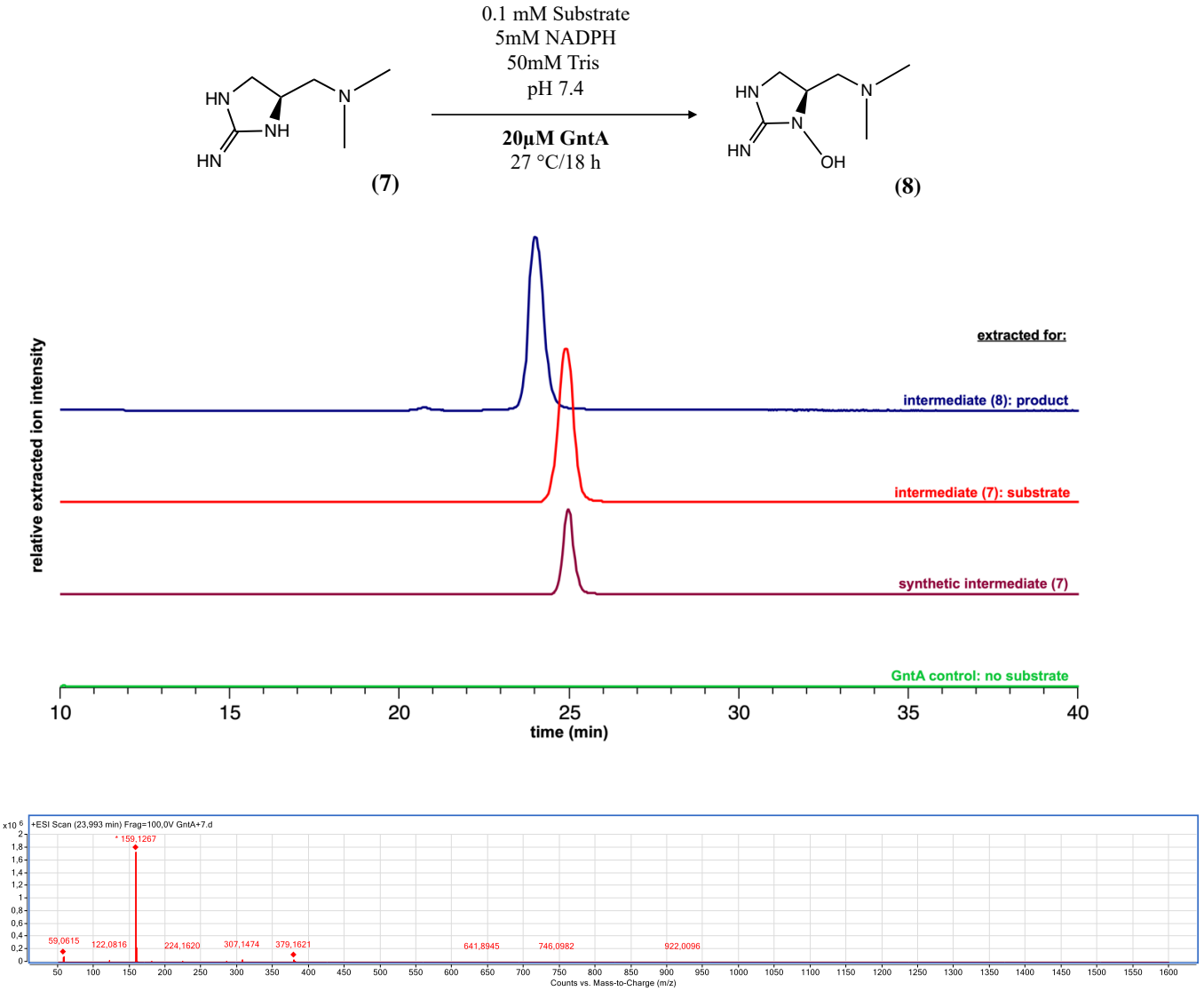
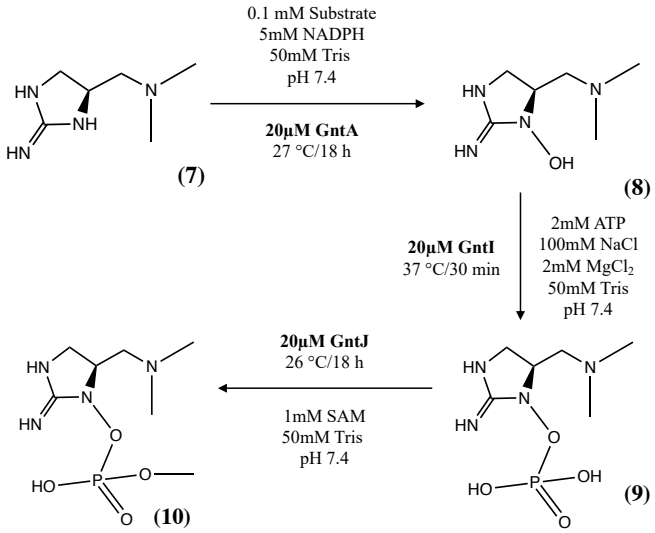
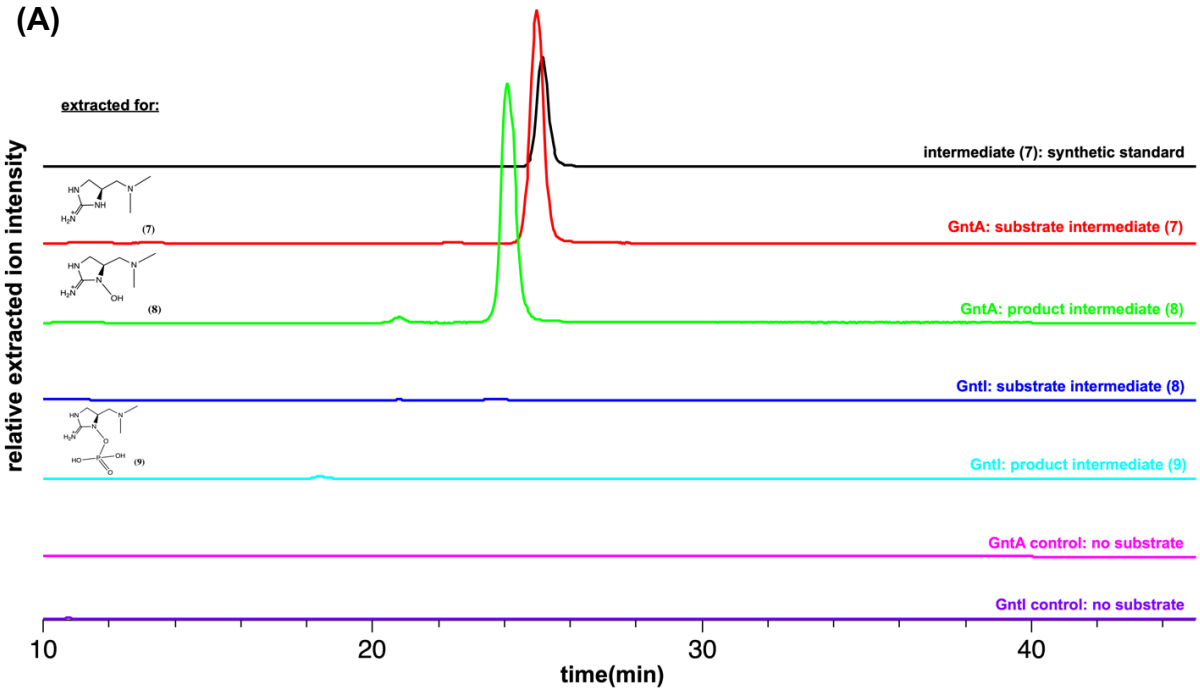


Fig. S10 - GntA reaction were set up as described before using 5 mM of NADPH and incubated for 18 hours at 27 ºC. Positive mode LCMS chromatograms using LC Method A, showing the extracted ion chromatogram for product, substrate, and synthetic substrate standard (EIC ± 0.01 *m/z*). Positive mode LCMS chromatogram of GntA reaction product shows *N*-hydroxylated product, (8), peak (+EIC 159.1267 ± 0.01 *m/z*, blue trace).


0.1 mM Substrate
5mM NADPH
50mM Tris
pH 7.4
20µM GntA
27 °C/18 h
(7)
(8)
20µM GntI
37 °C/30 min
2mM ATP
100mM NaCl
2mM MgCl2
50mM Tris
pH 7.4
20µM GntJ
26 °C/18 h
1mM SAM
50mM Tris
pH 7.4
(9)
(10)


(A)
extracted for:
intermediate (7): synthetic standard
GntA: substrate intermediate (7)
GntA: product intermediate (8)
GntI: substrate intermediate (8)
GntI: product intermediate (9)
GntA control: no substrate
GntI control: no substrate
relative extracted ion intensity
time(min)
10
20
30
40

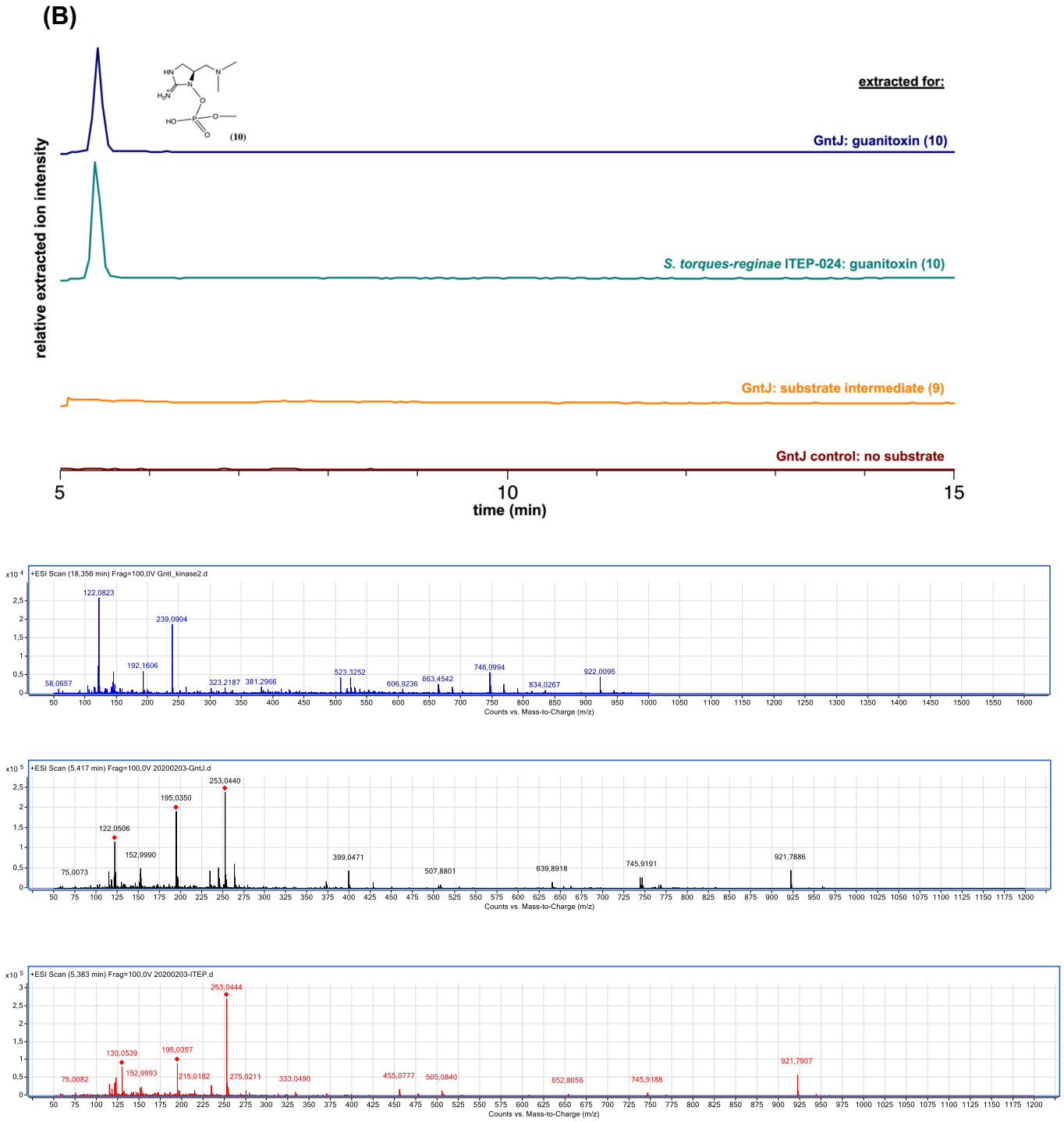


Fig. S11 - GntA, GntI and GntJ coupled reactions were set up as previous described starting from the (7) synthetic standard to guanitoxin (GNT). Positive mode LCMS chromatograms using LC Method A (A) for and Method B (B) show the extracted ion chromatogram for compounds (8), (9) and guanitoxin (10) (EIC ± 0.01 m/z). Phosphorylated intermediate (9) (light blue trace marked with a triangle (A)) (+EIC 239.0904 ± 0.01 m/z) and guanitoxin (dark blue trace (B)) (+EIC 253.1060 ± 0.01 m/z) production were confirmed *in vitro* using purified enzymes. Cyano trace (B) shows the guanitoxin peak (+EIC 253.1060 ± 0.01 m/z) in the cyanobacteria *S. torques-reginae* ITEP-024.

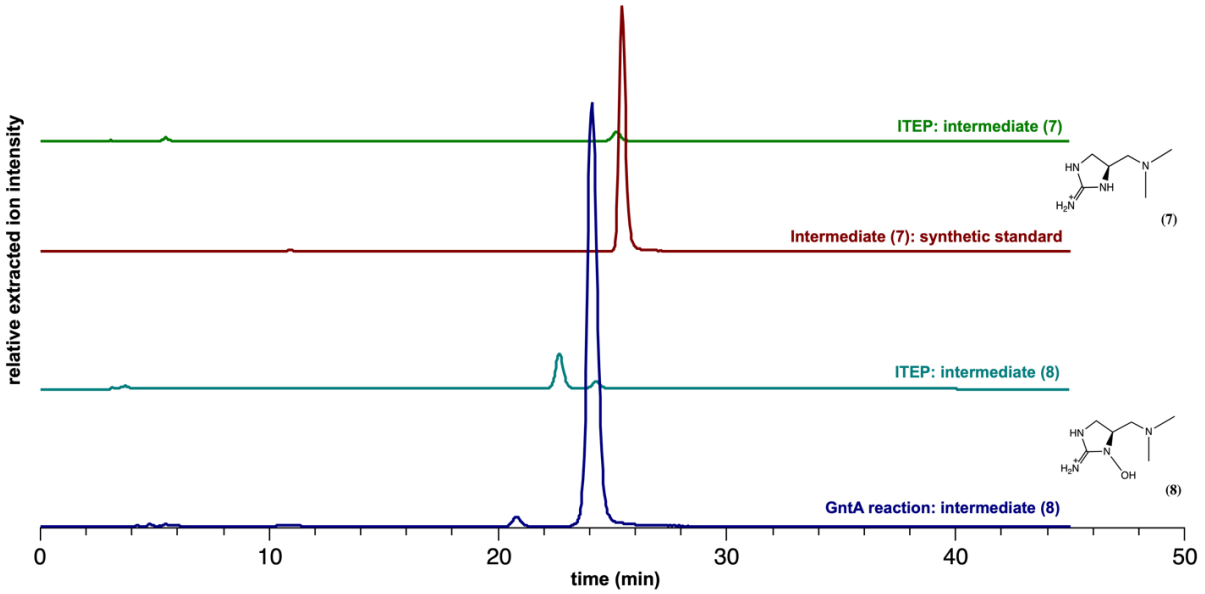


Fig. S12 - Positive mode LCMS chromatograms using LC Method A show the presence of (7) and (8) in the cyanobacteria culture (green trace and cyan trace marked with triangle, respectively), extracted as previous described. The peaks of cyanobacteria culture were compared with *in vitro* assays with GntA (N-hydroxylase) purified enzyme (dark red trace) and (7) standard, chemically synthetized (dark blue trace).

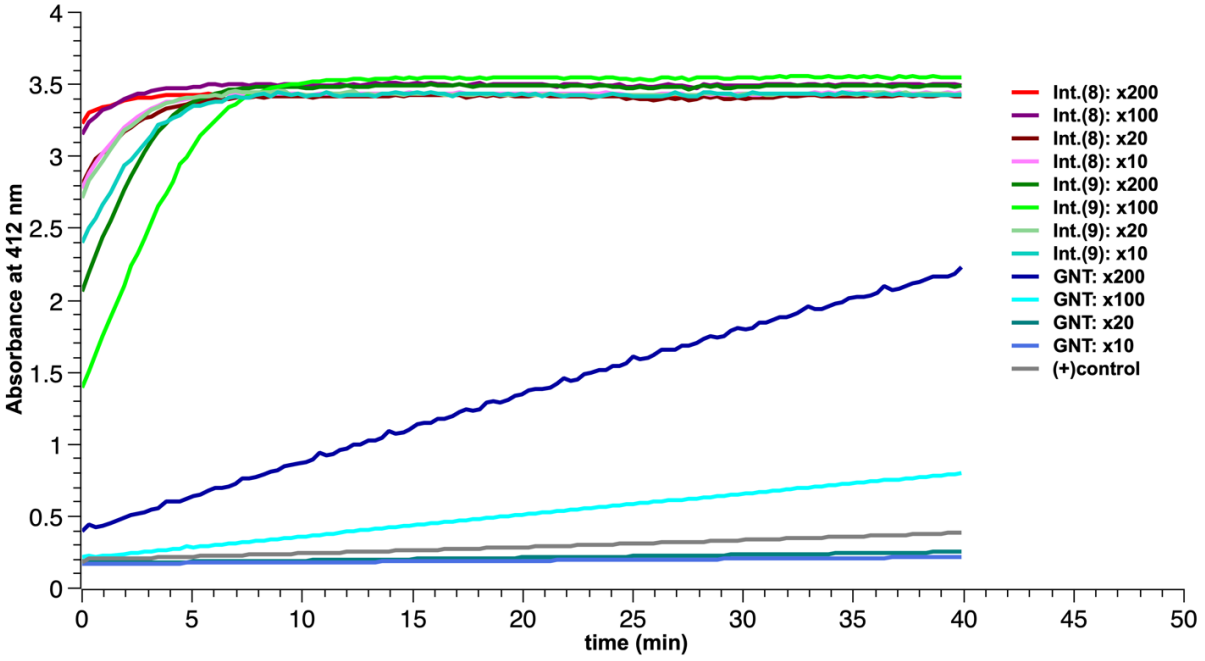
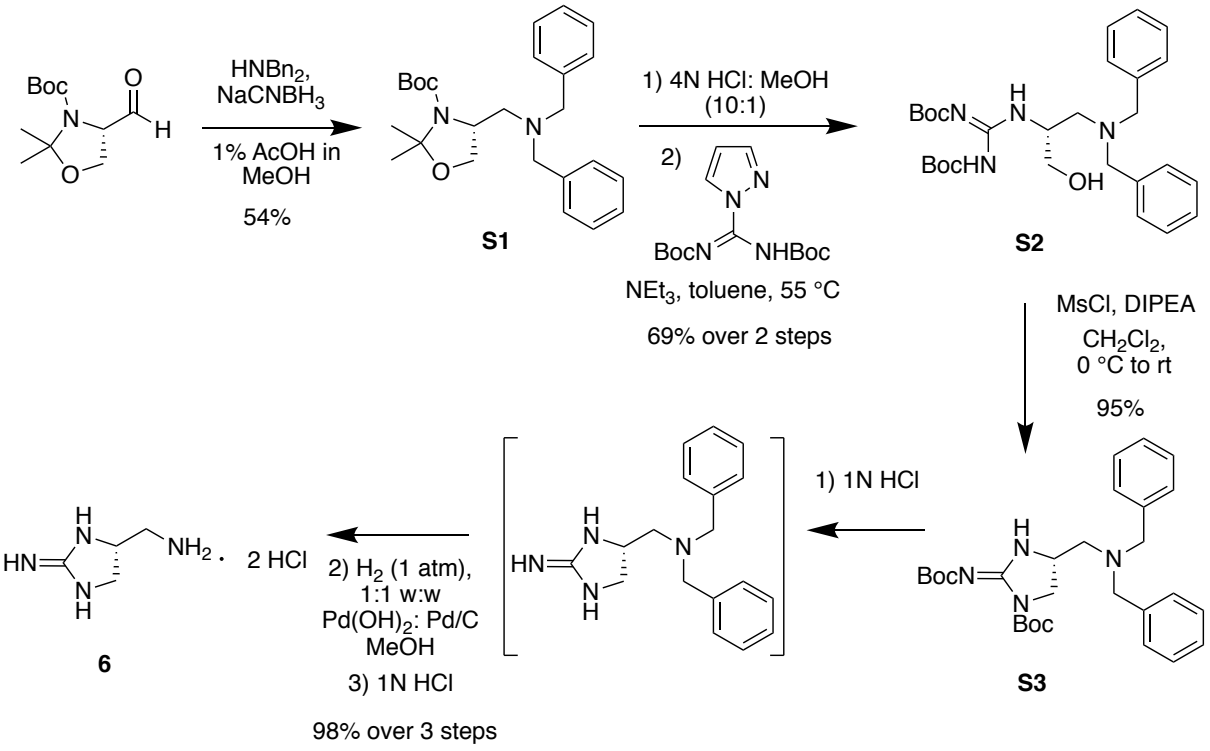


Fig. S13 - Acetylcholinesterase assay with (8), (9) and final product guanitoxin (GNT), under 200, 100, 20 and 10 times diluted. The reactions were carried out as described before using Bio Vision Acetylcholinesterase Inhibitor Screening Kit (Colorimetric). It utilizes the ability of an active human acetylcholinesterase (AChE) enzyme to hydrolyze the provided colorimetric substrate, generating a yellow chromophore that can be detected by measuring absorbance at 412 nm. In the presence of the potent reversible AChE inhibitor donepezil (positive control, gray trace), enzyme activity is suppressed, preventing chromophore generation. The assay shows that just after the addition of the methyl by GntJ (O-methyltransferase) was observed inhibition of AChE activity, pointing that 20 time and 10 time dilution (green and light blue traces, respectively) were more powerful than positive control, donepezil.

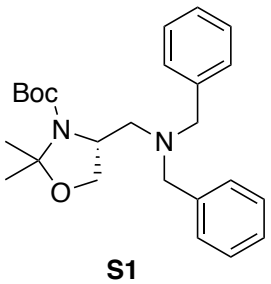
Chemical Synthesis

Two intermediates of GNT pathway were synthetized to be test with the correspondent enzymes, presumed to accept those compounds.

A) Primary amine compound: intermediate (**6**) of GNT pathway



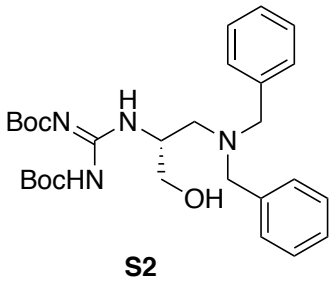
(*R*)-(2-iminoimidazolidin-4-yl)methanamine (**6**)



S1

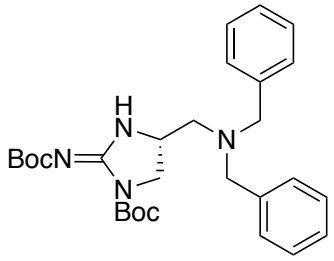
A solution of (S)-Garner's aldehyde (0.50 mL, 2.31 mmol) and dibenzylamine (0.47 mL, 2.43 mmol) in methanol (10 mL) had acetic acid (0.1 mL) added and was stirred for 15 minutes at room temperature. To this solution, sodium cyanoborohydride (0.145 g, 2.31 mmol) was added over a few minutes and stirred at room temperature under an argon atmosphere for 20 h. The reaction mixture was cooled to 0 °C and an aqueous solution of saturated sodium bicarbonate (25 mL) was added, followed by ethyl acetate (50 mL). The layers were separated, and the aqueous layer was further extracted with ethyl acetate (2 x 50 mL).

Pooled organic layers were washed with brine (50 mL), dried over magnesium sulfate, filtered and concentrated in vacuo. The crude reaction mixture was purified by silica flash chromatography and eluted over a gradient of 9:1 to 4:1 hexanes:ethyl acetate + 0.1% triethylamine. Pooled fractions were concentrated in vacuo, yielding the desired product as a clear light-yellow oil (0.514 g, 54%).



S2

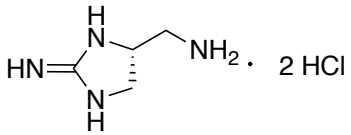
A solution of S1 (0.328 g, 0.80 mmol) in 4 N aqueous HCl (10 mL) and methanol (1 mL) was stirred at room temperature for 2 h. The reaction mixture was concentrated in vacuo, using water and toluene co-evaporations to remove additional HCl and water respectively. The crude material was resuspended in toluene (15 mL) and had triethylamine (0.56 mL, 3.99 mmol) and N,N'-di-Boc-1H-pyrazole-1-carboxamidine (0.273 g, 0.88 mmol) sequentially added. The reaction mixture was heated to 55 °C and stirred for 12 h, then was cooled to room temperature, diluted with ethyl acetate (50 mL) and washed with water (2 x 25 mL). The organic layer was dried over magnesium sulfate, filtered, and concentrated in vacuo. The crude reaction mixture was purified by silica flash chromatography and eluted over a gradient of 9:1 to 4:1 hexanes : ethyl acetate + 0.1% triethylamine. Pooled fractions were concentrated in vacuo, yielding the desired product as a sticky white foam (0.283 g, 69%).



S3

To a 0°C solution of S2 (0.091 g, 0.18 mmol) in dry CH2Cl2 (10 mL) was sequentially added DIPEA (62 μL, 0.36 mmol) and methanesulfonyl chloride (15 μL, 0.20 mmol). The reaction mixture was slowly warmed to room temperature over 21 h, then quenched by the addition of saturated NH4Cl (20 mL) and diluted with CH2Cl2 (10 mL). The layers were separated, and the organic layer was further washed with water (20 mL), and brine (20 mL), dried over magnesium sulfate, filtered and concentrated in vacuo. The crude reaction mixture

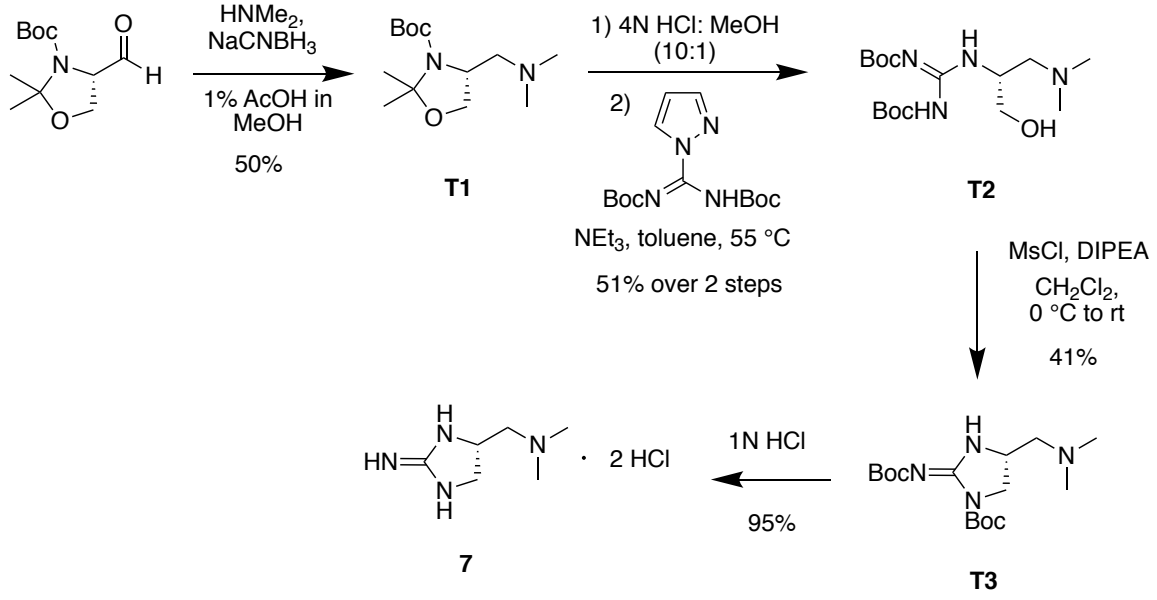
was purified by silica flash chromatography and eluted over a stepwise gradient of 99:1 to 49:1 chloroform:methanol + 0.1% triethylamine. Pooled fractions were concentrated in vacuo, yielding the desired product as a white solid (0.083 g, 95%).



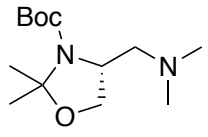
6

A solution of S3 (0.045 g, 0.091 mmol) in 1 N aqueous HCl (10 mL) was stirred at room temperature for 16 h, then concentrated in vacuo, using water and methanol co-evaporations to remove residual solvents. The crude reaction mixture was resuspended in methanol (10 mL) and had 20% Pd(OH)2 (15 mg, wet) and 10% Pd/C (30 mg) sequentially added. The reaction mixture was bubbled with hydrogen gas using a balloon (1 atm) and monitored by LCMS (C18 RP-HPLC) for consumption of the di-benzylated starting material (consumed after 30 minutes) and the mono-benzylated intermediate (consumed after 22 h overnight incubation). The reaction mixture was filtered through a pad of Celite, rinsed with methanol (30 mL), then 0.1% aqueous acetic acid (30 mL) and concentrated in vacuo. 1 N aqueous HCl (5 mL) was added to the filtrate to obtain the HCl salt, then concentrated in vacuo and lyophilized. The desired product was obtained as a light yellow solid.

B) <u>Dimethylamine compound: intermediate (7) of GNT pathway</u>

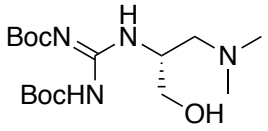


(*S*)-1-(2-iminoimidazolidin-4-yl)-*N*,*N*-dimethylmethanamine (**7**)



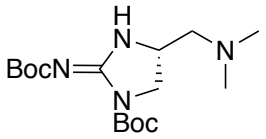
T1

A solution of (S)-Garner's aldehyde (0.560 g, 2.44 mmol) and dimethylamine (1.28 mL, [2.0 M solution in THF], 2.56 mmol) in methanol (10 mL) had acetic acid (0.1 mL) added and was stirred for 5 minutes at room temperature. To this solution, sodium cyanoborohydride (0.169 g, 2.69 mmol) was added over a few minutes and stirred at room temperature under an argon atmosphere for 24 h. The reaction mixture was cooled to 0 °C and an aqueous solution of saturated sodium bicarbonate (25 mL) was added, followed by ethyl acetate (50 mL). The layers were separated, and the aqueous layer was further extracted with ethyl acetate (2 x 50 mL). Pooled organic layers were washed with brine (50 mL), dried over magnesium sulfate, filtered and concentrated in vacuo. The crude reaction mixture was purified by silica flash chromatography and eluted over a gradient of 50:1 to 10:1 ethyl acetate:methanol + 0.1% triethylamine. Pooled fractions were concentrated in vacuo, yielding the desired product as a clear colorless oil (0.313 g, 50%).



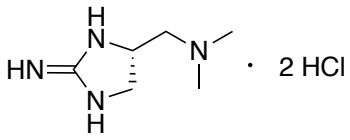
T2

A solution of T1 (0.245 g, 0.95 mmol) in 4 N aqueous HCl (10 mL) and methanol (1 mL) was stirred at room temperature for 2.5 h. The reaction mixture was concentrated *in vacuo*, using water and toluene co-evaporations to remove additional HCl and water respectively. The crude material was resuspended in toluene (15 mL) and had triethylamine (0.66 mL, 4.74 mmol) and *N*,*N'*-di-Boc-1*H*-pyrazole-1-carboxamidine (0.324 g, 1.04 mmol) sequentially added. The reaction mixture was heated to 55 °C and stirred for 17 h, then was cooled to room temperature, diluted with ethyl acetate (50 mL) and washed with water (2 x 25 mL). The organic layer was dried over magnesium sulfate, filtered, and concentrated *in vacuo*. The crude reaction mixture was purified by silica flash chromatography and eluted over a stepwise gradient of 50:1 to 33:1 to 20:1 to 9:1 ethyl acetate : methanol + 0.1% triethylamine. Pooled fractions were concentrated *in vacuo*, yielding the desired product as a light-yellow oil (0.175 g, 51%).



T3

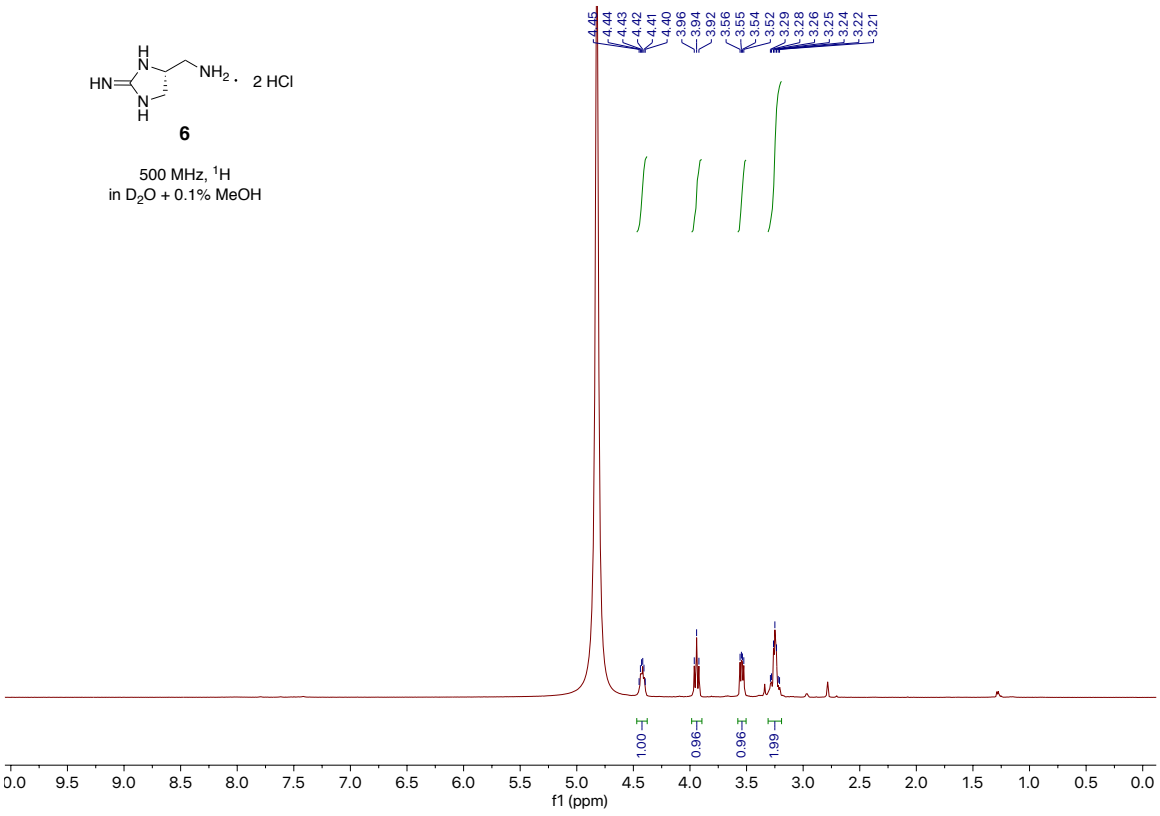
To a 0°C solution of T2 (0.122 g, 0.34 mmol) in dry CH2Cl2 (10 mL) was sequentially added DIPEA (118 μL, 0.68 mmol) and methanesulfonyl chloride (29 μL, 0.37 mmol). The reaction mixture was slowly warmed to room temperature over 18 h, then quenched by the addition of saturated NH4Cl (20 mL) and diluted with CH2Cl2 (10 mL). The layers were separated, and the organic layer was further washed with water (20 mL), and brine (20 mL), dried over magnesium sulfate, filtered and concentrated in vacuo. The crude reaction mixture was purified by silica flash chromatography and eluted over a gradient of 20:1 to 9:1 chloroform:methanol + 0.1% triethylamine. Pooled fractions were concentrated in vacuo, yielding the desired product as a white solid (0.047 g, 41%).

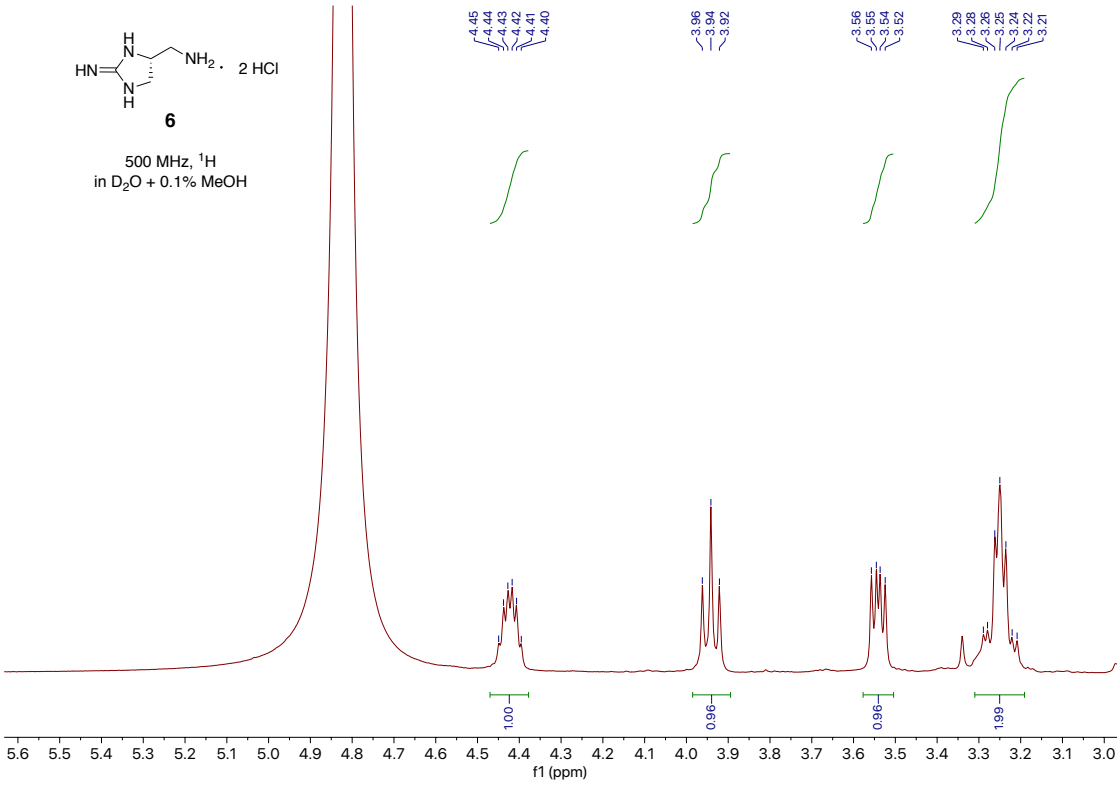


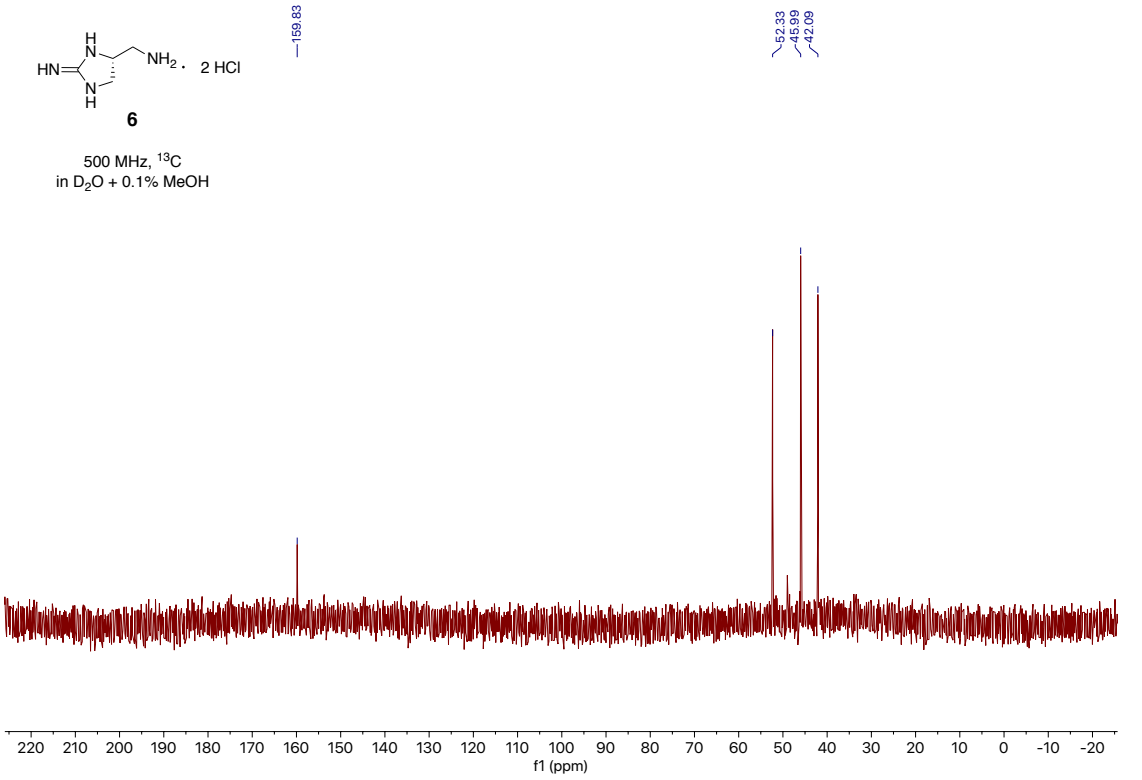
7

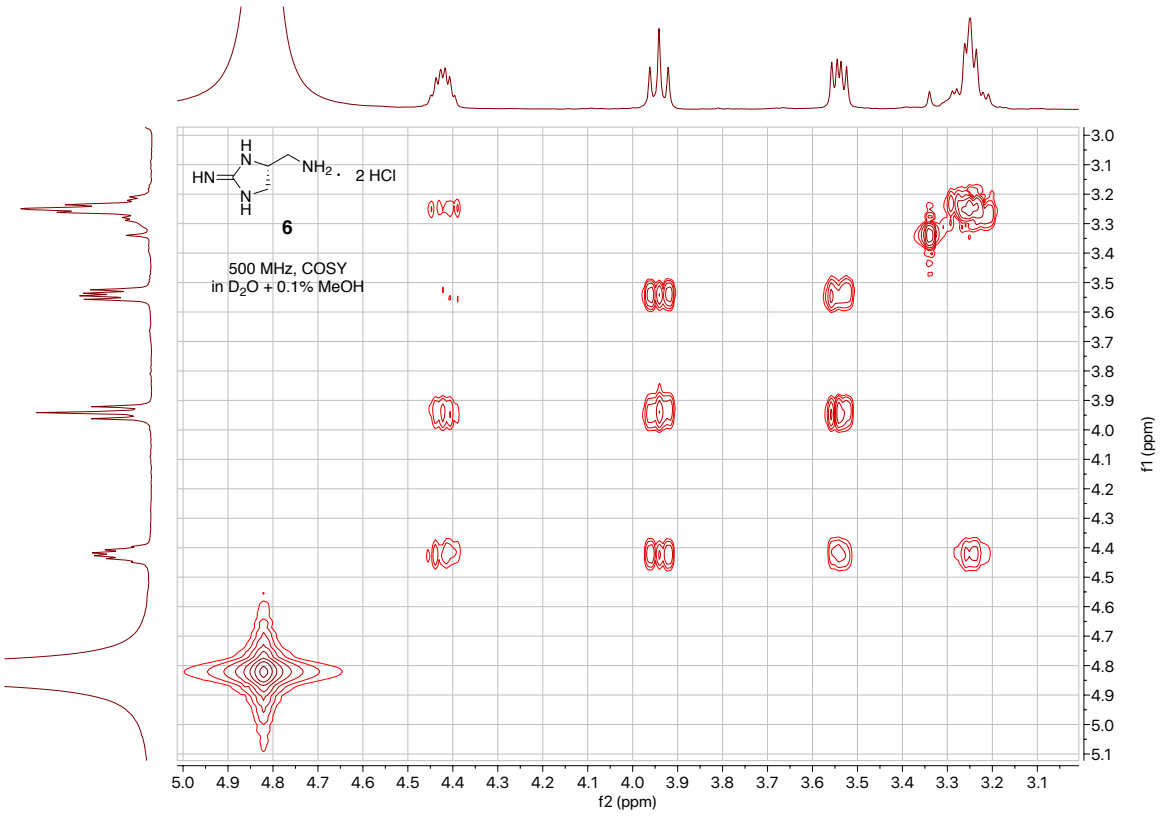
A solution of T3 (0.032 g, 0.093 mmol) in 1N aqueous HCl (5 mL) was stirred at room temperature for 14 h, then concentrated *in vacuo* and lyophilized. The desired product was obtained as a light yellow solid (0.019 g, 95%).

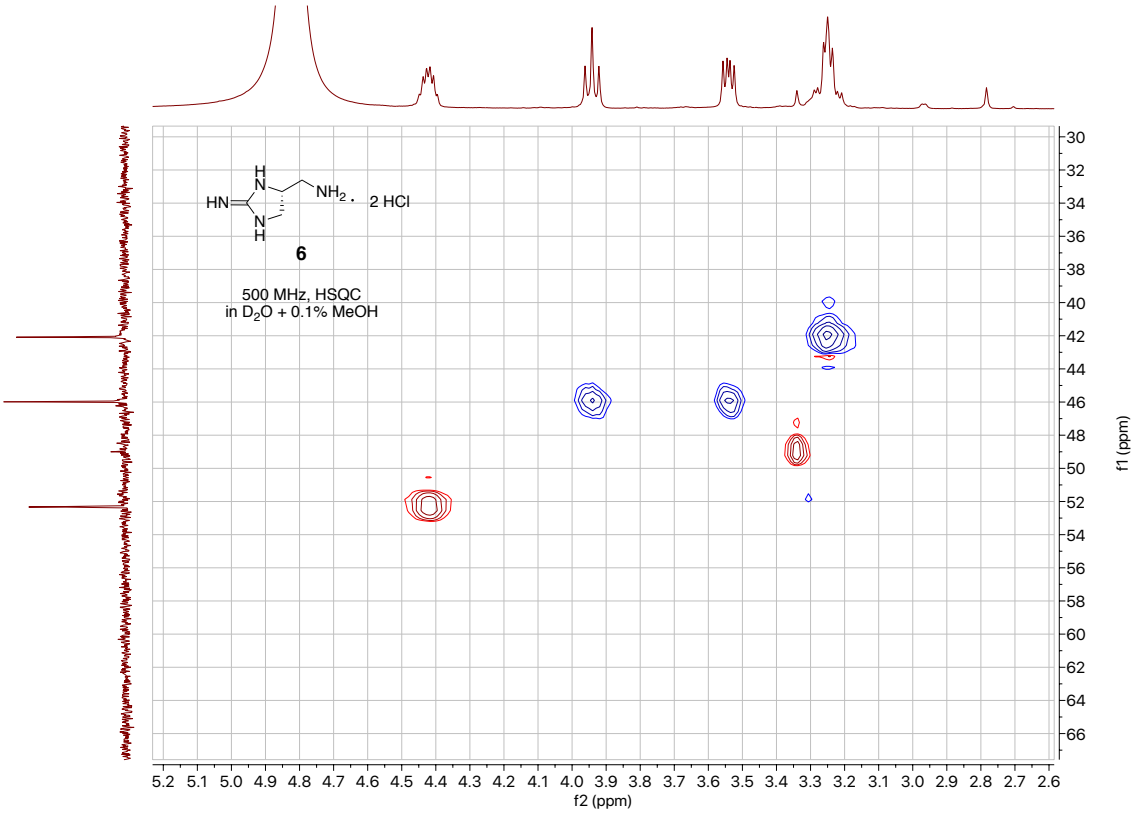
C) NMR spectras

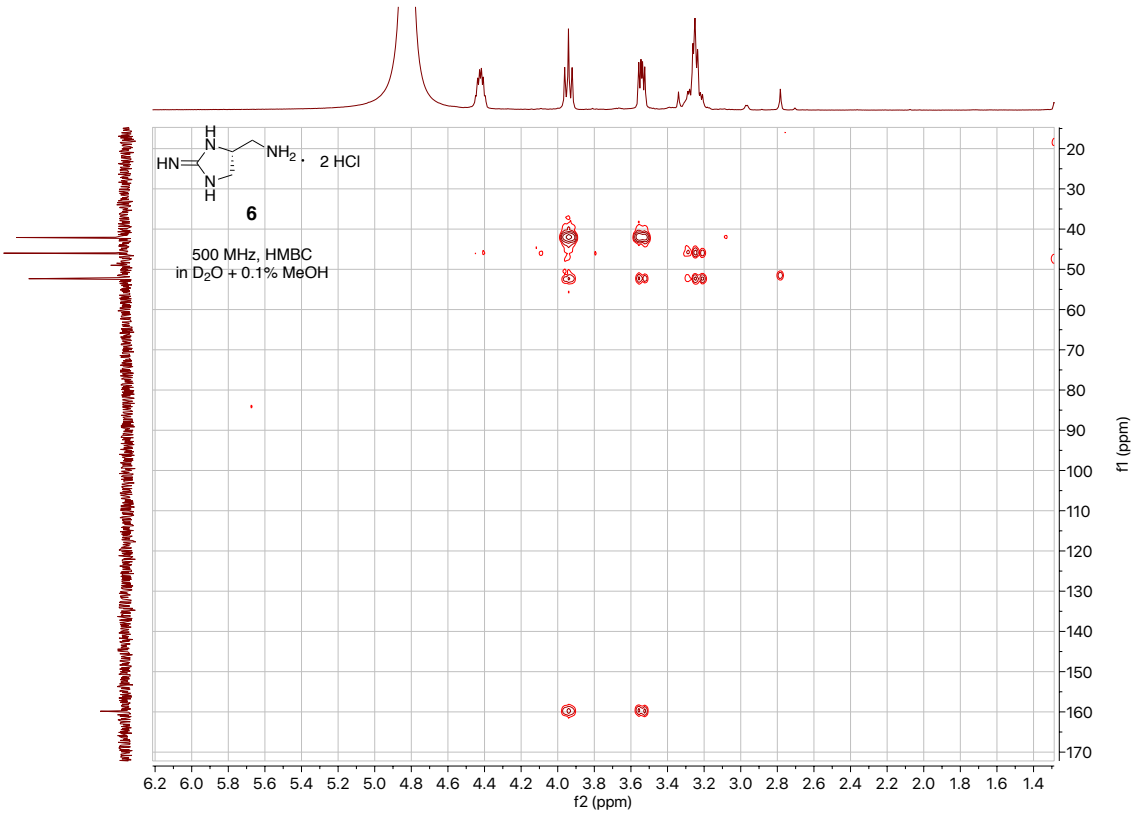


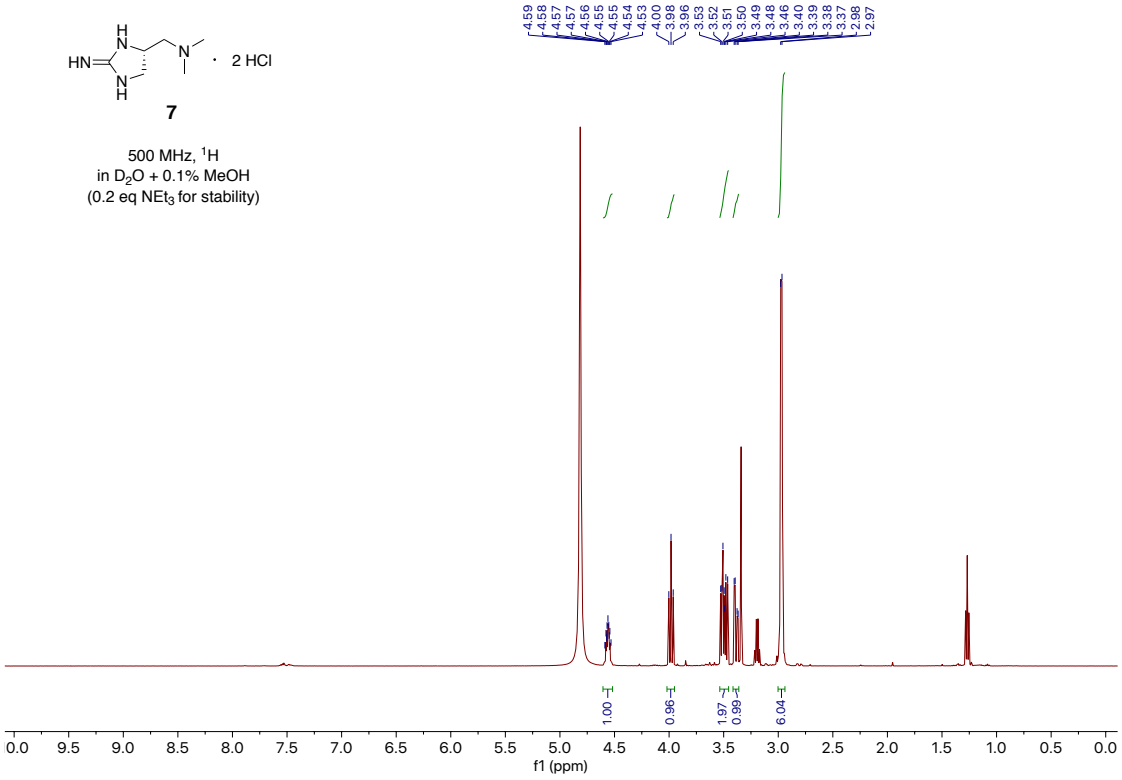


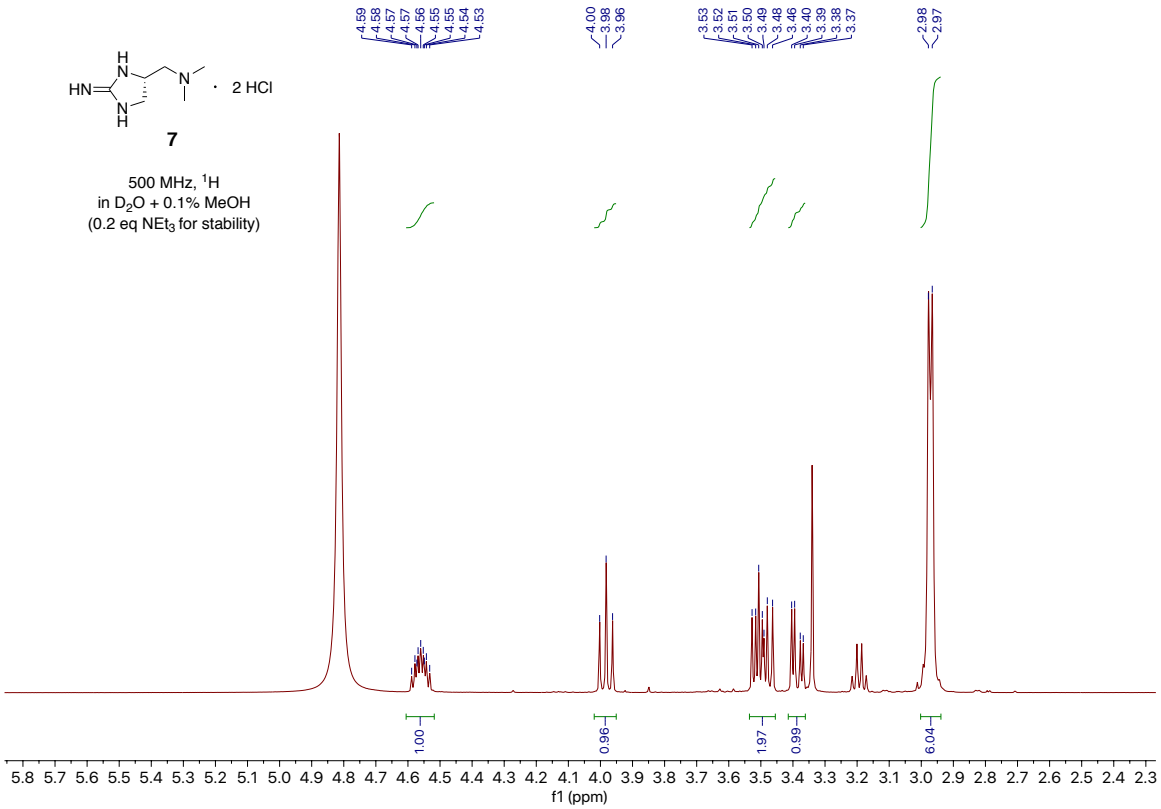

159.83
52.33
45.99
42.09
NH2 . 2 HCl
6
500 MHz, 13C
in D2O + 0.1% MeOH
f1 (ppm)

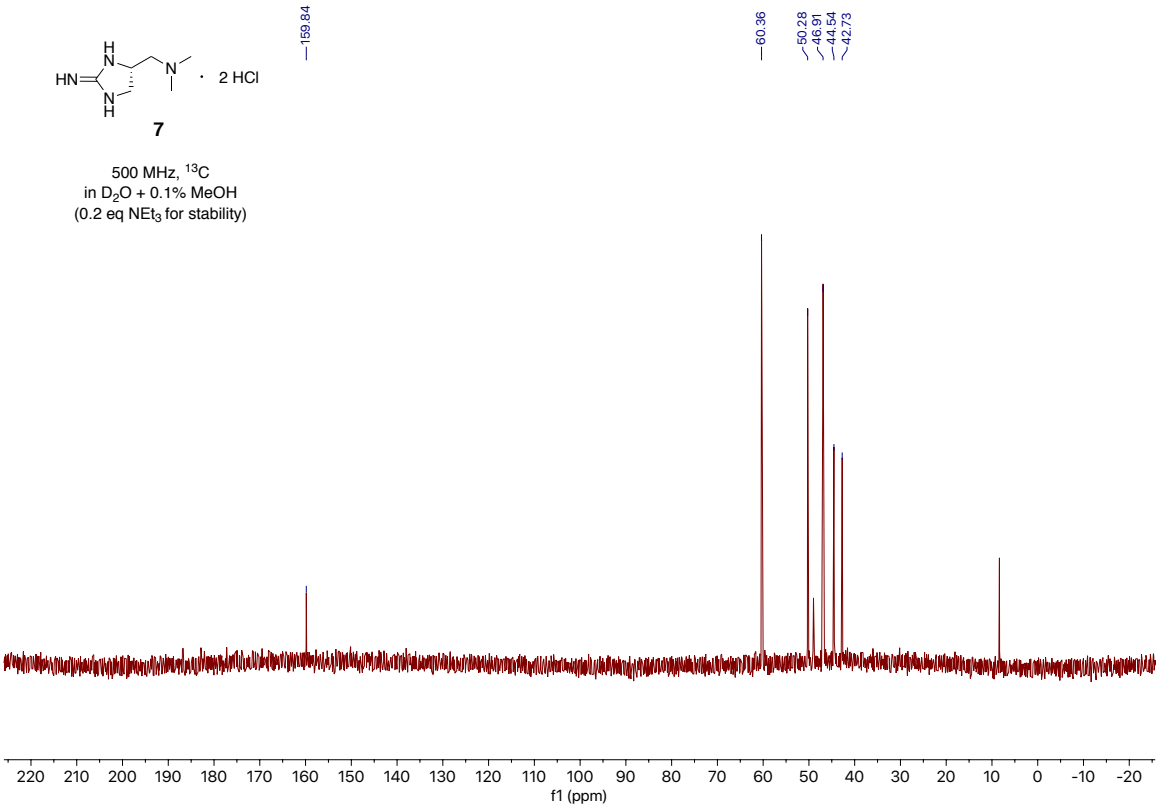

NH2 . 2 HCl
6
500 MHz, COSY
in D2O + 0.1% MeOH
f2 (ppm)
f1 (ppm)

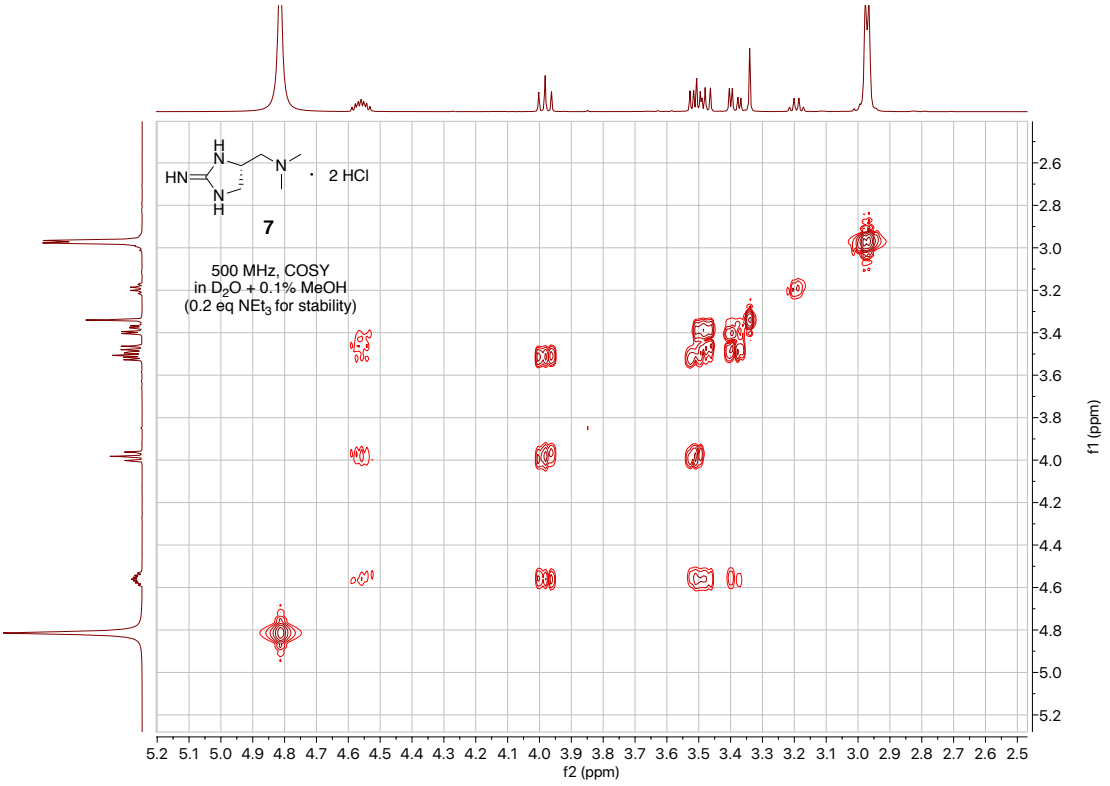

6
500 MHz, HSQC
in D2O + 0.1% MeOH
2 HCl
f2 (ppm)
f1 (ppm)

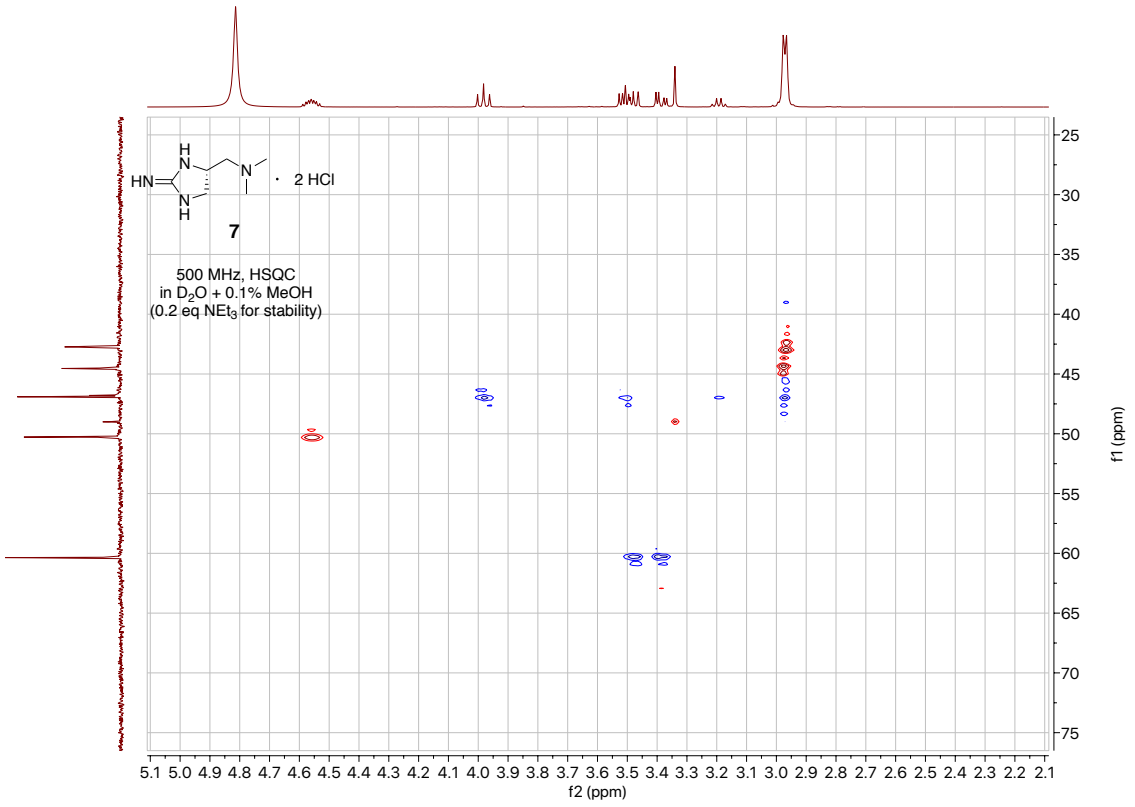

6
500 MHz, HMBC
in D2O + 0.1% MeOH
2 HCl
f2 (ppm)
f1 (ppm)

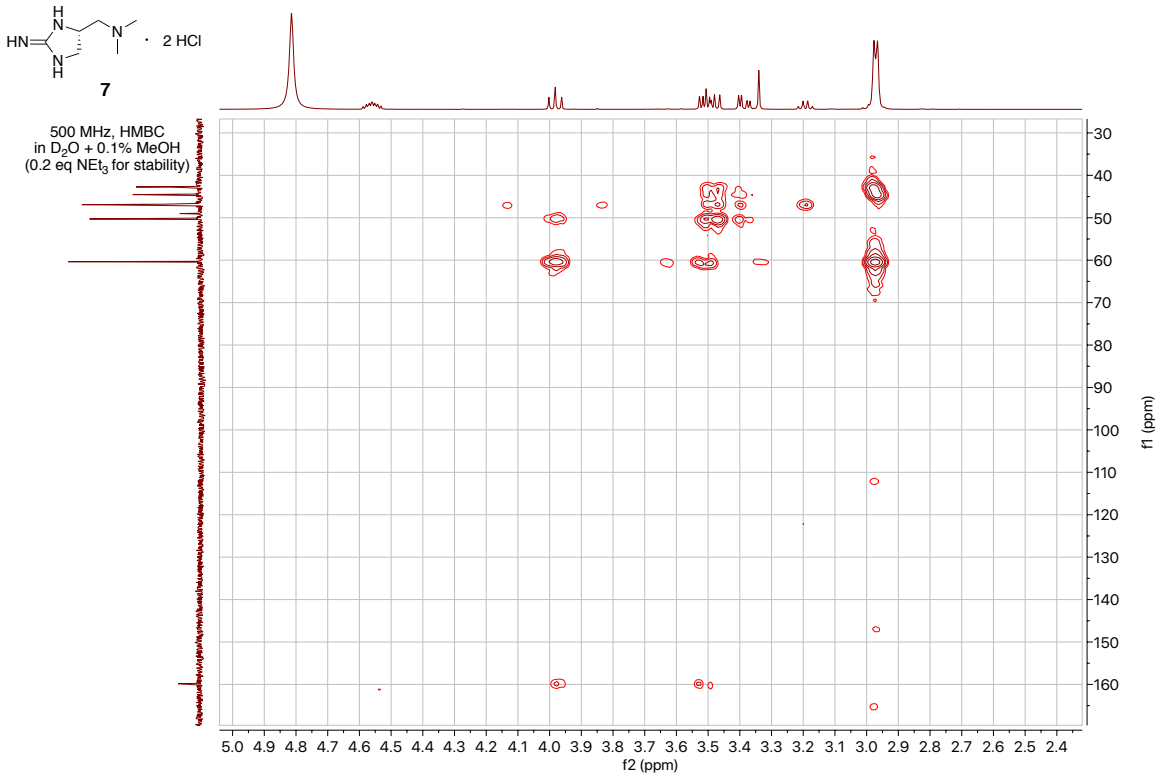

· 2 HCl
7
500 MHz, ^{1}H
in D_2O + 0.1% MeOH
(0.2 eq NEt_3 for stability)
4.59 4.58 4.57 4.57 4.56 4.55 4.55 4.54 4.53
4.00 3.98 3.96
3.53 3.52 3.51 3.50 3.49 3.48 3.46 3.40 3.39 3.38 3.37
2.98 2.97
1.00 0.96 1.97 0.99 6.04
f1 (ppm)


· 2 HCl
7
500 MHz, ^{1}H
in D_2O + 0.1% MeOH
(0.2 eq NEt_3 for stability)
4.59 4.58 4.57 4.57 4.56 4.55 4.55 4.54 4.53
4.00 3.98 3.96
3.53 3.52 3.51 3.50 3.49 3.48 3.46 3.40 3.39 3.38 3.37
2.98 2.97
1.00 0.96 1.97 0.99 6.04
f1 (ppm)


159.84
60.36
50.28
46.91
44.54
42.73
2 HCl
7
500 MHz, 13C
in D2O + 0.1% MeOH
(0.2 eq NEt3 for stability)
f1 (ppm)


2 HCl
7
500 MHz, COSY
in D2O + 0.1% MeOH
(0.2 eq NEt3 for stability)
f1 (ppm)
f2 (ppm)


HN
N
N
H
H
· 2 HCl
7
500 MHz, HSQC
in D2O + 0.1% MeOH
(0.2 eq NEt3 for stability)
f2 (ppm)
f1 (ppm)


HN
N
N
H
H
· 2 HCl
7
500 MHz, HMBC
in D2O + 0.1% MeOH
(0.2 eq NEt3 for stability)
f2 (ppm)
f1 (ppm)

Supplementary Table

Table S1 - Primer table using in this study.

gntB-F	TTGTTTAACTTTAATAAGGAGATATACCATGAAGAAAAACATCAAGAAA TACCGTTTC
gntB-R	GCATTATGCGGCCGCAAGCTTGTTAGCTGCTAACTTGGGTCAGA
gntC-F	GTTAAGTATAAGAAGGAGATATACATATGAAGATCCAGCCGGCGCTG
gntC-R	CCGATATCCAATTGAGATCTGCCATATGTTAGCTGGTTTCAATAACGATC GCCAG
pCOLA-F	CAAGCTTGCGGCCGCATAATGC
pCOLA-R	CATGGTATATCTCCTTATTAAAGTTAAACAA

References

DE ROND, T.; STOW, P.; EIGI, I.; JOHNSON, R.E.; CHAN, L.J.G.; GOYAL, G.; BAIDOO, E.E.K.; HILLSON, N.T.; PETZOLD, C.J.; SARPONG, R.; KEASLING, J.D. Oxidative cyclization of prodigiosin by an alkylglycerol monooxygenase-like enzyme. **Nature Chemical Biology**. New York, v. 13, n. 11, p. 1155-1157, 2017.

MOLICA, R.J.R.; OLIVEIRA, E.J.A.; CARVALHO, P.V.V.C.; COSTA, A.N.S.F.; CUNHA, M.C.C.; MELO, G.L.; AZEVEDO, S.M.F.O. Occurrence of saxitoxins and an anatoxin-a(s)-like anticholinesterase in a Brazilian drinking water supply. **Harmful Algae**. Amsterdam, v. 4, n. 4, p. 743–753, 2005.

PARKS D.H.; CHUVOCHINA, M.; WAITE, D.W.; RINKE, C.; SKARCHWSKI, A.; CHAUMEIL, P.A.; HUGENHOLTZ, P. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. **Nature Biotechnology**. New York, v. 36, p. 996-1004, 2018.

Appendix B: Crystal structure of GntC

We obtained one X-ray crystallographic structure of GntC to a resolution of 2.1Å in complex with PLP (Figure 1). This structure crystallized as a dimer of dimers in space group $P2_1$. OrfR (PDB: 4M2M) was used as a model for molecular replacement. The GntC consist of four monomers (370 residues) (Figure 2) and PLP is bounded to Lys219 residue (Figure 3 and 4). Unfortunately, it was not possible to crystallize the enzyme and substrate together since the substrate for this enzyme has not been commercially available.

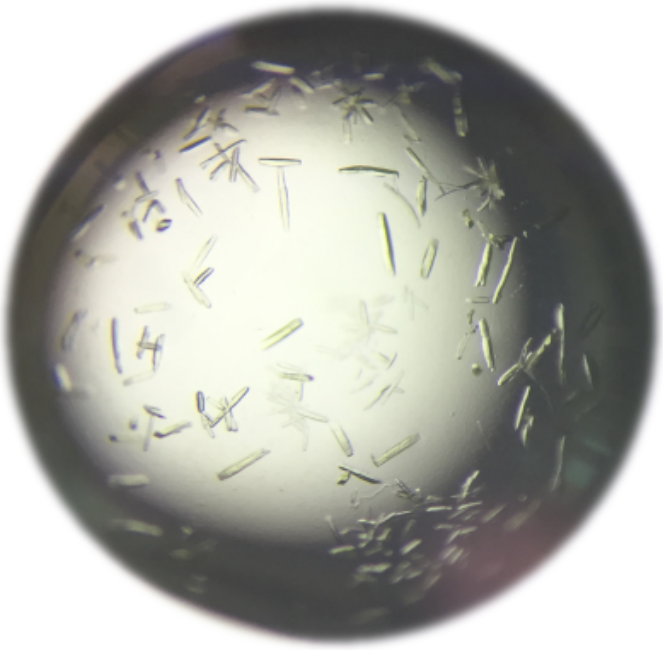


Figure 1 – GntC drop shows the protein crystals in PEG 3,350, $MgCl_2$, 0.1 M Tris pH 8.5 and 1 mM PLP. In the image is possible to see a light-yellow color in the crystals, which indicate the PLP bind in the protein.

Table 1 – Refinement statistics for crystallography data.

	GntC + PLP
Data collection	
Space group	$P2_1$
Cell dimensions	
a, b, c (Å)	60.85, 158.16, 73.23
α, β, γ (°)	90.00, 90.01, 90.00
Resolution (Å)	73.2 - 2.10 (2.106 - 2.099)
R_{sym} (%)	12.7 (80.2)
R_{pim} (%)	5.4 (37.2)
$I / \sigma I$	10.9 (2.2)
Completeness (%)	93.1 (97.3)

Redundancy	5.2 (4.6)
$CC_{1/2}$	0.99 (0.57)
Refinement	
Resolution (Å)	73.2 - 2.1
No. reflections	75056
R_{work} / R_{free} (%)	21.4 / 25.6
No. atoms	
Protein	11014
Water	381
B-factors (Å^2)	
Protein	33.5
Water	34.0
R.m.s. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	1.055

Highest resolution shell is shown in parenthesis. R-factor = Σ ($|F_{obs}|$-k$|F_{calc}|$)/Σ $|F_{obs}|$ and R-free is the R value for a test set of reflections consisting of a random 5% of the diffraction data not used in refinement.

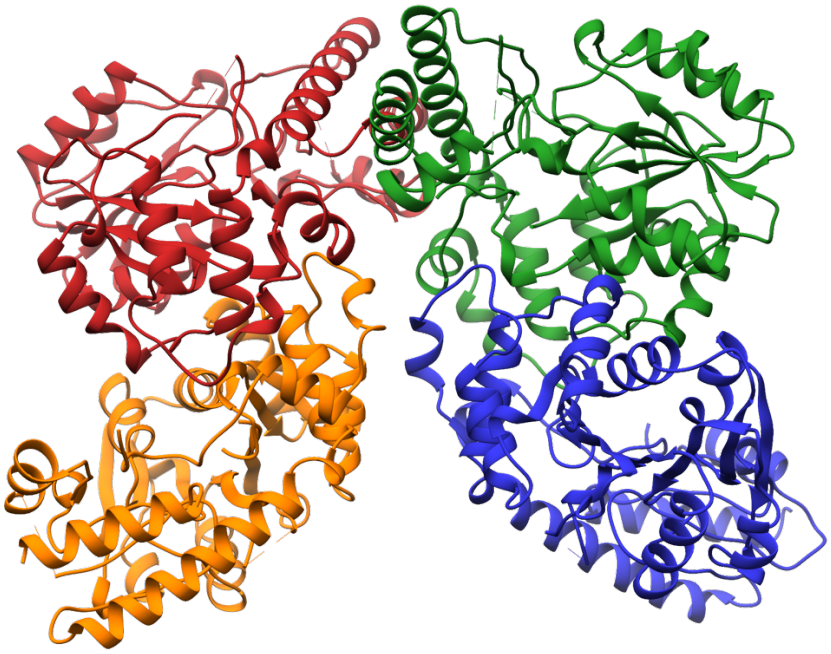


Figure 2 - GntC dimer. The colors indicate the four monomers that constitute the protein.

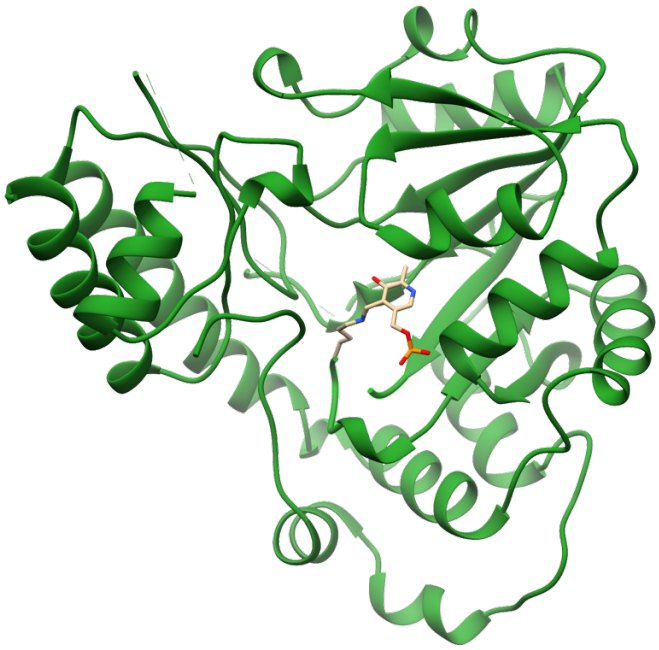


Figure 3 - Monomer chain showing the PLP bound.

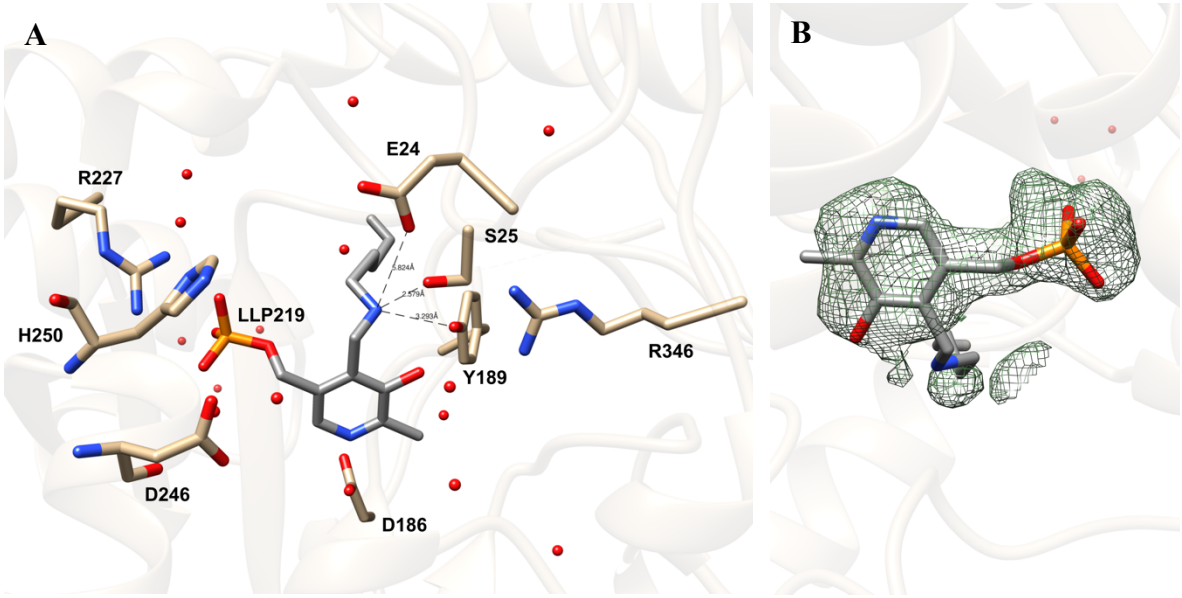


Figure 4 – **(A)** Active site of GntC showing the PLP bound. **(B)** F_o-F_c maps were generated by removing the PLP and completing refinement without the cofactor. Mesh was contoured to 2σ (green). This electron density map indicates the presence of PLP covalently bound to Lys219 in the active site.

The high structural homology of GntC and OrfR is consistent with that GntC is acting as an unusual PLP-dependent enzyme to cyclize an oxidized arginine intermediate and form enduracididine as the second step in the guanitoxin biosynthetic pathway. Figure 5 shows the homology between GntC and OrfR, which catalyzes the PLP-dependent elimination/addition

reaction cyclizing (3*R*,4*R*)-(OH)2-L-Arg to the six-membered (4*R*)-OH-capreomycidine in the streptolidine biosynthetic pathway (CHANG et al., 2014).

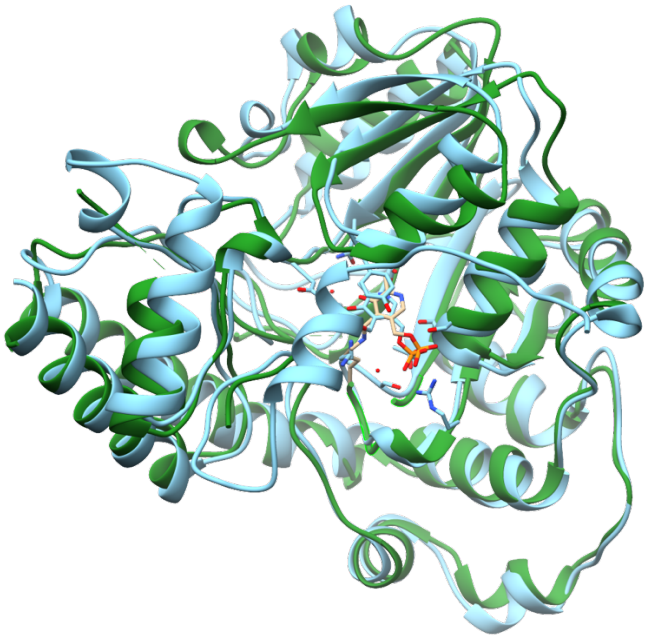


Figure 5 - Protein structural homology between GntC (green) and OrfR (blue).

Based on the homology with OrfR, a cyclization reaction has been proposed for GntC, the PLP-dependent cyclodehydratase (Figure 6). First the 4(*S*)-hydroxy-L-arginine binds to PLP to form an aldimine adduct. The active site base then removes the Cα proton of the substrate and the PLP conjugation system provides formation of a double bound between Cα and Nα. A second active site base deprotonates Cβ, forming a double bond between Cα and Cβ, with Nα serving as an electron sync. This intermediate can then undergo dehydration and re-formation of a double bond between Cα and Nα. This enables attack of guanidine group by the terminal amine to form the five-member ring of guanitoxin molecule. Finally, hydrolyses take place in the aldimine linkage to free L-enduracididine (Figure 6).

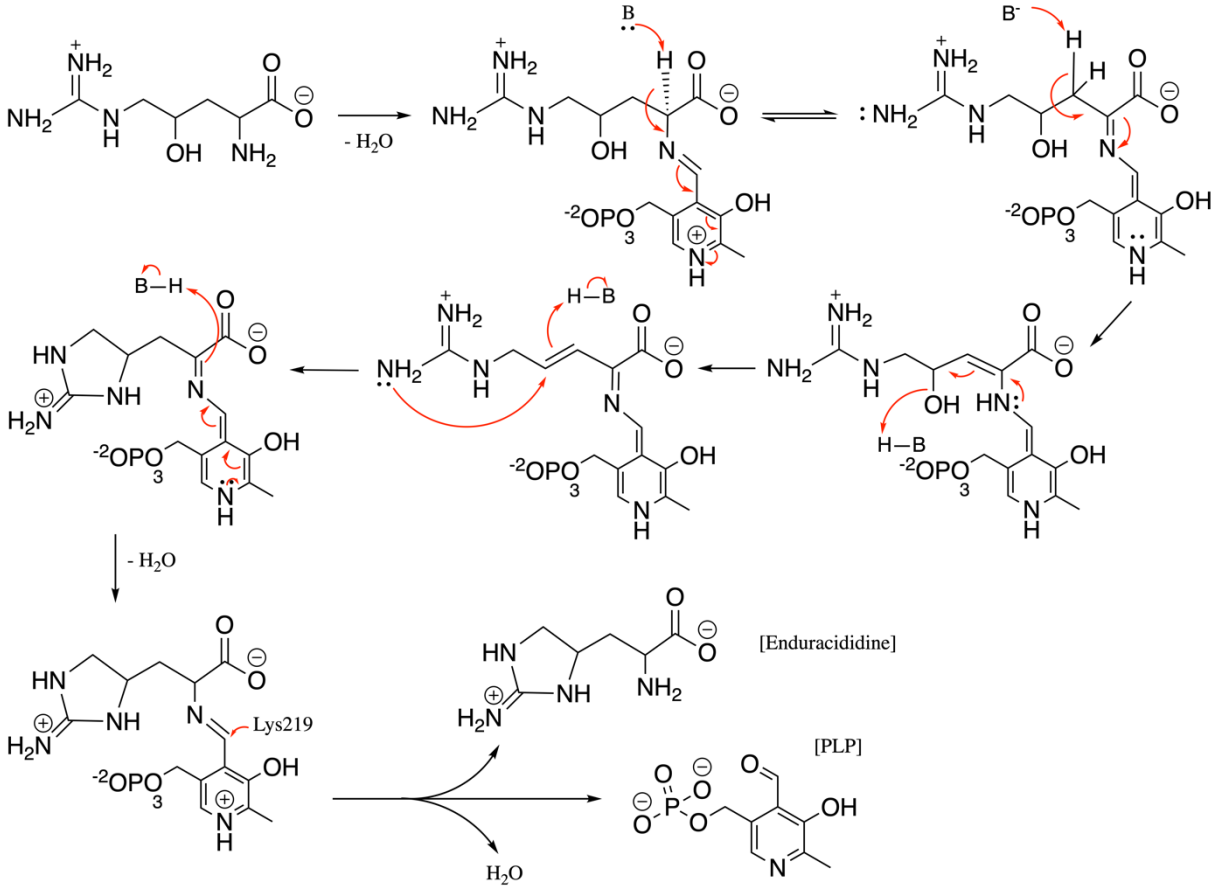


Figure 6 - Proposed mechanism for cyclodehydration reaction to generate enduracididine by GntC (PLP-dependent cyclodehydratase). The red arrows show the electrons pushing.

References

CHANG, C.Y.; LYU, S.Y; LIU, Y.C; HSU, N.S; WU, C.C.; TANG, C.F.; LIN, K.H.; HO, J.Y.; WU, C.J.; TSAI, M.D.; LI, T.L. Biosynthesis of streptolidine involved two unexpected intermediates produced by a dihydroxylase and a cyclase through unusual mechanisms. **Angewandte Chemie International Edition**. New York, v. 53, p. 1943 –1948, 2014.

Appendix C: Article published during the doctoral regarding results achieved in the master's degree

Genetic organization of anabaenopeptin and spumigin biosynthetic gene clusters in the cyanobacterium *Sphaerospermopsis torques-reginae* ITEP-024[2]

Stella T. Lima,[†] Danillo O. Alvarenga,[†] Augusto Etchegaray,[‡] David P. Fewer,[§] Jouni Jokela,[§] Alessandro M. Varani,[•] Miriam Sanz,[‖] Felipe A. Dörr,[‖] Ernani Pinto,[‖] Kaarina Sivonen,[§] Marli F. Fiore[†*]

[†]University of São Paulo, Center for Nuclear Energy in Agriculture, Piracicaba-SP, Brazil
[‡]Pontifical Catholic University of Campinas, Faculty of Chemistry, Campinas-SP, Brazil
[§]University of Helsinki, Department of Food and Environmental Sciences, Division of Microbiology and Biotechnology, Helsinki, Finland
[•]São Paulo State University, College of Agricultural and Veterinary Sciences, Department of Technology, Jaboticabal-SP, Brazil
[‖]University of São Paulo, School of Pharmaceutical Sciences, São Paulo-SP, Brazil

* Corresponding author: Marli Fátima Fiore, University of São Paulo, Center for Nuclear Energy in Agriculture. E-mail: fiore@cena.usp.br

ACS Chemical Biology: Volume 12, March 2017, 769-778.

Keywords: Genomics. Protease inhibitors. Non-Ribosomal Peptides. Natural products. Chemotypes.

Abstract

Cyanobacteria produce a broad range of natural products many of which are potent protease inhibitors. Biosynthetic gene clusters encoding the production of novel protease inhibitors belonging to the spumigin and anabaenopeptin family of non-ribosomal peptides were identified in the genome of the bloom-forming cyanobacterium *Sphaerospermopsis torques reginae* ITEP-024. The genetic architecture and gene organization of both non-ribosomal

[2] Article accepted 13 January 2017. Published 17 March 2017

peptide biosynthetic clusters were compared in parallel with their chemical structure variations obtained by liquid chromatography (LC-MS/MS). The spumigin (*spu*) and anabaenopeptin (*apt*) gene clusters are co-located in the genomes of *S. torques-reginae* ITEP-024 and *Nodularia spumigena* CCY9414 and separated by a 12 kb region containing genes encoding a patatin-like phospholipase, L-homophenylalanine (L-Hph) biosynthetic enzymes and four hypothetical proteins. *hphABCD* gene cluster encoding the production of L-Hph was linked to all eight apt gene clusters investigated here. We suggest that while the HphABCD enzymes are an integral part of the anabaenopeptin biosynthetic pathway, they provide substrates for the biosynthesis of both anabaenopeptins and spumigins. The organization of the *spu* and *apt* suggests a plausible model for the biosynthesis of the 4-(4-hydroxyphenyl)-2-acid (Hpoba) precursor of spumigin variants in *S. torques-reginae* ITEP-024 based on the acceptable substrates of HphABCD enzymes.

Introduction

Cyanobacteria encode an impressive array of natural products many of which are made using nonribosomal peptide synthetases (NRPS).[1–3] Many of these molecules possess pharmacological or biological activity with potential application as drug candidates.[4–8] Natural product discovery has been accelerated by advances in cyanobacterial genome sequencing efforts and genome mining approaches.[9–12] Protease inhibitors play an important role in regulating cell survival, host defense, development and tissue homeostasis.[13] Therefore, cyanobacterial peptides with protease inhibitory activity are of particular interest in potential clinical applications against hypertension, myocardial infarction, periodontitis, AIDS, thrombosis, respiratory disease pancreatitis and cancer.[14–18]

The description of new variants of natural products and the elucidation of their respective biosynthetic genes is valuable for pharmaceutical development, since structural variants often have different biological activities.[3,19] Spumigins and anabaenopeptins are both protease inhibitors produced by cyanobacteria through NRPS biosynthetic pathways.[20,21] Spumigins are linear tetrapeptides characterized by the presence of an N-terminal hydroxyphenyl lactic acid, variable hydrophobic amino acid, proline or methyl proline and arginine (or lysine) derivatives.[20,22,23] These protease inhibitors were described for the first time in the strain *Nodularia spumigena* AV1.[22] Twenty-four spumigin isoforms have been reported so far in strains of the genera Nodularia, Anabaena and Sphaerospermopsis.[20,22–25] However, the only gene cluster currently available is from the strain *N. spumigena* CCY9414.[20]

Anabaenopeptins are a diverse class of cyclic hexapeptides characterized by the presence of D lysine and an unusual ureido linkage.[2] The discovery and structural elucidation of anabaenopeptins were first reported in the strain *Anabaena flos-aquae* NRC 525-17.[26] The nomenclature of this class of protease inhibitors is exceedingly complex and chemical variants of anabaenopeptins are named according to producing organisms, strain source, place of isolation, or molecular weight, amongst others.[27] Thus, they can be ferintoic acids, nodulapeptins, oscillamides, schizopeptin, nostamide, pompanopeptin or lyngbyaureidamides amongst others.27 There are reports of 115 isoforms of anabaenopeptins, of which 104 are produced by cyanobacteria, eight have been found from theonellid sponge, one from sponge *Psammocinia* and two from an oyster.[27, 28-33] The macrocyclic structure of anabaenopeptin contains five amino acid residues connected by an ureido bond to an exocyclic amino acid.[34,35] The general structure of anabaenopeptin can be represented as X1-CO-[Lys-X3-X4-X5-X6], wherein the X3-6 are variable non-proteinogenic or proteinogenic amino acids of the cycle. The cyclization bond is formed by the C-terminal carboxyl with the N-terminal amine of lysine.[21] The ureido linkage and D-lysine configuration are conserved, whereby the other residues have varied configurations.[27] Anabaenopeptins are known from strains of the genera *Anabaena*, *Aphanizomenon*, *Microcystis*, *Nodularia*, *Nostoc*, *Planktothrix*, *Oscillatoria*, *Schizothrix* and *Lyngbya*.21,22,26,36–42 The anabaenopeptin biosynthetic gene cluster (*apt*) was firstly identified from Planktothrix rubescens NIVA-CYA 98.[39] The apt gene cluster has been characterized in strains from the genera *Anabaena*, *Nostoc*, *Nodularia* and *Planktothrix*.[21,34,39,43] Recent investigation found apt gene clusters accompanied by three biosynthetic genes (*hphABCD*), which catalyze the biosynthesis of homophenylalanine (L-Hph).[43] L-Hph is a non-proteinogenic amino acid and useful chiral building block for the synthesis of several pharmaceutical drugs.[43]

The genus *Sphaerospermopsis* comprises planktonic, filamentous heterocytous cyanobacteria of the order Nostocales and family Aphanizomenonaceae.[44] *Sphaerospermopsis torques-reginae* can proliferate at high densities mainly in brackish and freshwater, form cyanobacterial blooms, and is widely distributed in South America.[45] Previous chemical analysis of *S. torques-reginae* ITEP-024 cell extracts demonstrated the production of the neurotoxin anatoxin-a(s)[46] and novel spumigin variants.[25] Here, we characterized a unique nucleotide region in the sequenced genome of strain *S. torques-reginae* ITEP-024 containing spumigin and anabaenopeptin biosynthetic gene clusters. The close physical proximity of the *spu* and *apt* gene clusters identified in this study most likely reflects an overlap among spumigin and

anabaenopeptin biosynthetic pathways allowing the provision of substrates for the biosynthesis of both non-ribosomal peptides.

Results and Discussion

Here we describe a genomic region found in the genome of *S. torques-reginae* ITEP-024, which encodes *spu* and *apt* gene clusters. A 22.9 kb spumigin biosynthetic gene cluster and a 23.4 kb anabaenopeptin gene cluster were identified on a single contig in the draft genome of *S. torques reginae* ITEP-024. A more detailed analysis of the upstream and downstream boundaries of the *spu* and *apt* biosynthetic gene clusters revealed that these clusters are separated by a 12 kb region, which encodes a patatin-like phospholipase, L-homophenylalanine (L-Hph) biosynthetic enzymes and four hypothetical proteins (Figure 1). An almost identical gene organization was identified after mining the *N. spumigena* CCY9414 genome, which produces both spumigins and anabaenopeptins.[20,21] The anabaenopeptins produced by ITEP-024 and CCY9414 strains contain homophenylalanine, whereas only spumigins produced by CCY9414 contain homotyrosine.[20,21] The hph biosynthetic genes are encoded in the apt gene cluster suggesting that they may supply homophenylalanine or homotyrosine for the synthesis of both protease inhibitors. In all genomes of oscillatorialean, chroococcalean and nostocalean strains evaluated the *hphABCD* gene cluster appear linked to the apt gene cluster (Figure 1). It was hypothesized that HphA was involved in making the ureido linkage in the biosynthesis of anabaenopeptins in nostocalean strains, and thus this gene was named *aptE*.[21] However, biochemical evidence demonstrates that the HphABCD proteins are involved in making L-Hph[43], and the *aptE* gene of nostocalean strains in fact belongs to the *hphABCD* gene set. The *hphA* gene has been annotated as benzylmalate synthase and participates in the conversion of phenylpyruvate into homophenylalanine.[43] The biosynthetic origins of the ureido moiety at N-terminal leucine in anabaenopeptin biosynthesis remain unclear.[21] Therefore, we have renamed the *aptF* gene, which encodes an ATP binding cassette (ABC) transporter, to *aptE* (Figure 1), so the nomenclature in this study is in agreement with both the apt gene clusters found in *Planktothrix rubescens* NIVA-CYA 98[39] and *Planktothrix agardhii* NIVA-CYA 126/8.[34]

HphA accepts both phenylpyruvic acid and 4-hydroxylphenylpyruvate (Hpp) substrates and the HphABCD enzymes can produce L-Hph and also L-homotyrosine from L-tyrosine.[43] Homotyrosine and homophenylalanine are both present in some spumigin and anabaenopeptin variants but their biosynthetic origins are unclear.[20,21] The co-location of the *spu* and *apt* gene clusters in *S. torques-reginae* ITEP-024 and *N. spumigena* CCY9414 suggests that these peptides would likely share the homophenylalanine and/or homotyrosine produced by the

HphABCD enzymes. A previous study mentioned that the gene clusters for nodularin, spumigin, nodulapeptin and aeruginosin were not randomly distributed but cluster in a 0.8 Mb region of the *N. spumigena* CCY9414 genome.[47] In order to test if the co-location of genes clusters encoding these protease inhibitors is a common feature in cyanobacteria, additional spumigin and anabaenopeptin producer genomes must become available.

The *apt* gene cluster of *S. torques-reginae* ITEP-024 has a similar structural organization compared to the apt gene clusters from *Anabaena*, *Nostoc*, *Nodularia* and *Planktothrix*.[21,34,39,43] Minor differences were observed in the presence of the additional gene (ORF1), encoding an hypothetical protein, between *aptB* and *aptC* genes and the nucleotide sequences of each gene. ORF1 did not show identity to any previously known protein sequence, and thus it is not possible to predict its biosynthetic role. The nucleotide sequence of each apt gene showed identities varying from 80% to 93% to apt gene sequences found in the anabaenopeptin-producing *N. spumigena* CCY9414 (Supplementary Table S1). The *apt* gene cluster codes four NRPS enzymes, AptA, AptB, AptC, and AptD, containing altogether six NRPS modules, each of which having an adenylation domain (Figure 2). AptA (248 kDa) is the first protein encoded in the apt gene cluster harboring two modules composed by two adenylation (A) domains, two peptidyl carrier protein (PCP) domains and one condensation (C) domain. AptB (122 kDa) includes C, A, and PCP domains, organized in a single module. AptC (292 kDa) is composed by two modules containing two C domains, two A domains, two PCP domains and one methyltransferase (NMT) domain. Finally, AptD (159 kDa) is composed by A, C, PCP, and thioesterase (TE) domains, organized in a single module. The TE domain is responsible for releasing and cyclizing the anabaenopeptin from the mega-synthetase. AptE (91 kDa) is an ATP binding cassette (ABC) transporter.

Phylogenetic reconstruction based on concatenated nucleotide sequences of representative complete apt gene clusters from all known producers of anabaenopeptin placed the *S. torques-reginae* ITEP-024 *apt* gene cluster together with that of *N. spumigena* CCY9414 (Figure 3). Furthermore, the nostocalean strains clustered together in a separated clade from the cluster formed by the oscillatorialean and chroococcalean strains (Figure 3). The phylogenetic tree, based on the *apt* genes is almost identical to the accepted taxonomic position of the producing cyanobacteria, which suggests vertical transmission of *apt* gene clusters (Figure 3).

The anabaenopeptin variants produced by the strain S. torques-reginae ITEP-024 were predicted based on the six AptA-A1, AptA-A2, AptB-A, AptC-A1, AptC-A2, AptD-A NRPS adenylation domain binding pockets deduced from the gene cluster (Table 1). Substrate specificity predictions indicated the incorporation of isoleucine, lysine, valine, leucine, alanine,

and phenylalanine, according to the A domain selectivity found in the apt gene cluster. However, leucine has never been described in position 4 from anabaenopeptins originating directly from cyanobacteria. Unsurprisingly these amino acids predicted by the *in silico* analysis did not totally correlate with those obtained by mass spectrometry analysis of the strain ITEP-024 intracellular peptides (Figure 2, Supplementary Figure S1 and Table S3). The main differences were observed at positions 3 and 4. The amino acids at position 3 could be either valine or leucine/isoleucine based on in silico deduction and LC-MS/MS analyses, respectively. At position 4, leucine could be incorporated according to in silico prediction, but LC-MS/MS analyses confirmed the incorporation of leucine or homotyrosine. However, empirical molecular formula and fragmentation behavior were in accordance with the anabaenopeptin *m/z* 808 (Supplementary Figure S1) previously described in several strains of N. spumigena isolated from coastal water bodies of Southern Australia.[24] Anabaenopeptin *m/z* 808 contains lysine at position 2 as other anabaenopeptins, homotyrosine at position 4 (only Hph/Hty or their derivatives have been previously described in position 4), a N-methylated residue (methylalanine) at position 5 and phenylalanine at position 6. Mass spectrometry could not confirm which leucine isomer (leucine or isoleucine) was incorporated at positions 1 and 3.

The *spu* gene cluster of S. torques-reginae ITEP-024 contains six genes encoding enzymes responsible for the assembly of the linear tetrapeptide spumigin. This gene cluster contains an additional ORF1 between *spuD* and *spuE* (Figure 1) compared to the *spu* gene cluster of *N. spumigena* CCY9414. Detailed BLAST, UniProt, and Pfam analyses of the corresponding ORF1 (22 kDa) protein sequence indicated that this gene encodes a CoA-dependent acetyltransferase belonging to the maltose O-acetyltransferase (MAT)-like. However, the spumigins produced by *S. torques-reginae* ITEP-024 did not contain an acetyl group. Detailed analysis of product ions from protonated spumigins, proton- and 13C-HSQC NMR spectra could not demonstrate the presence of O-, N- or C-acetyl group in the spumigin structures (Supplementary Figure S2 and Table S4). Therefore, it is possible that this gene is silent or perhaps plays a role in the biosynthesis of precursor amino acids.

The *spu* cluster encodes two NRPS mega-synthetases, SpuA and SpuB, which together encode four modules (Figure 4). SpuA (158 kDa), is a hybrid NRPS/PKS containing the A, ketoreductase (KR), and PCP domains. The *S. torques-reginae* ITEP-024 SpuA is homologous to an AMP-dependent synthetase from *Aphanizomenon flos-aquae* NIES-81 (Supplementary Table S2). The second protein, SpuB (466 kDa), includes three NRPS modules composed of eleven domains (Figure 4). The first SpuB module contains C, A, PCP, and epimerization (E) domains; the second module contains C, A, and PCP domains; and the third one contains C, A,

PCP, and thioester reductase (TD) domains. SpuB is homologous to NRPS of the *N. spumigena* CCY9414 (Supplementary Table S2). Genetic architecture and gene organization of the SpuA and SpuB proteins are consistent with the order of the reactions proposed as necessary to assemble the tetrapeptide (Figure 4). SpuC (31 kDa) is a hypothetical protein homologous to another hypothetical protein from *N. spumigena* CCY9414 (Supplementary Table S2). SpuC detailed analyses using BLAST and the Uniprot database revealed homology to proteins acting on the amino acid proline, which indicate that SpuC may in fact have specificity for this substrate. SpuD (40 kDa) is a dehydrogenase homolog to a zinc-dependent dehydrogenase of *N. spumigena* CCY9414 and SpuE (29 kDa) is a reductase homolog to the NADP-dependent reductase of *N. spumigena* CCY9414 (Supplementary Table S2). SpuC, SpuD and SpuE are involved in the assembly line through the biosynthesis of the rare, non-proteinogenic amino acid (2S,4S)-4-methylproline, which is found in some spumigin variants.[20,22–25] SpuF (80 kDa) is an ATPbinding cassette (ABC) transporter, which shows high similarity to a member of a ABCtransporter transmembrane region from *N. spumigena* CCY9414 (Supplementary Table S2). This protein may be involved in the efflux of spumigin from the cell.

The first step in the assembly of the spumigin tetrapeptide is the recognition and incorporation of the precursor substrate by the SpuA-A domain. Results from NRPSpredictor2 analyses for SpuA-A suggest the potential incorporation of the non-proteinogenic amino acid α-hydroxyisocaproic acid, which seems to be the closest aliphatic hydroxy acid present in these databases (Table 2). This could be explained because the α-hydroxy-isocaproic acid is similar to a portion of 4-hydroxyphenylpyruvic acid (Hpp) and 2-hydroxy-4-(4-hydroxyphenyl) butyric acid (Hhpba) structures, respectively the pyruvic and butyric acid portions, which may not be present in the databases. In the strain *N. spumigena* CCY9414 SpuA-A domain activates Hpp whose C2 carbonyl group is reduced to hydroxyl group by the KR domain generating Hpla.[20] These authors also noticed that SpuA-A domain can directly activate Hpla. Previous LC-MS/MS analysis of the S. torques-reginae ITEP-024 strain revealed the presence of four spumigin isoforms containing Hhpba[25], which was confirmed by NRM (personal communication, E. Pinto, USP). Hhpba differs from Hpla in one additional methylene group and the biosynthetic origins of Hhpba are unclear. However, bioinformatics analysis allows us to propose a plausible model. SpuA-A domain could activate 4-(4-hydroxyphenyl)-2-oxobutanoic acid (Hpoba) directly, followed by C2 carbonyl is reduced to a hydroxyl by the KR domain and incorporated as Hhpba into spumigin (Figure 4) as found by mass spectrometry.[25] It is possible that Hpoba is supplied to SpuA-A through action of the HphABCD enzymes (Figure 5). Interestingly, these enzymes also accept phenylpyruvic acid

and 4-hydroxylphenyl pyruvic acid as substrates.[43] Therefore, as in the synthesis of homophenylalanine and homotyrosine, these enzymes would potentially be involved on the addition of one carbon to Hpp, producing Hpoba. This suggests that the *spu* and *apt* biosynthetic gene clusters encode a versatile set of enzymes for supplying both aromatic homoamino acids and unusual keto acid precursors to spumigin and anabaenopeptin NRPS enzymes. Further investigations are needed in order to elucidate the precise interaction among these mega enzymatic complexes.

The substrates for SpuB predicted by NRPSpredictor2 were tyrosine, proline and arginine (Table 2), corroborating the tandem mass spectrometry results.[25] This also confirmed the incorporation of tyrosine at position 2, proline or methylproline in position 3 and argininol or argininal in position 4. The latter are respectively the alcohol and aldehyde derivatives of arginine, both predicted to be incorporated given the domain architecture of the synthetase, which ends with a TE domain. The additional genes *spuC*, *spuD*, *spuE*, which are located downstream to the *spuB* gene, are potentially involved in the biosynthesis of the rare and nonproteinogenic amino acid 4-methylproline.[20]

Experimental Procedures

Cyanobacterial strain and growth conditions

The cyanobacterium strain ITEP-024 was isolated from a *Sphaerospermopsis torques-reginae* bloom sample collected in March 27th, 2002 in the Tapacurá. reservoir located in the municipality of Recife, Pernambuco, Brazil (8º02'14''S, 35º09'46''W) as previously described.48 This strain was kindly provided by Dr. Vera R. Werner from the Natural Sciences Museum, Zoobotanical Foundation of Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil, and is maintained in the culture collection of CENA/USP, Piracicaba, SP, Brazil, in ASM-1 medium49 under 14/10 h dark/light cycle with white fluorescent light (55 μmol photon m^{-2} s^{-1}) at 22 ± 1 ºC.

Total genomic DNA extraction

Cyanobacterium cells were grown in 125 mL Erlenmeyer flasks containing 50 mL of ASM-1medium for 20 days at the same condition described above. The 50 mL culture was centrifuged for 10 min at 7,690 xg and the cells were washed with ultrapure water, 0.05% Extran, washing solution (50 mM NaCl, 10 mM Tris-HCl pH 7.5, 2.5 mM EDTA pH 8.0 and 50% ethanol), and saline solution. All washes were followed by homogenization and centrifugation for 10 min at gravitational forces ranging from 2,307 xg to 7,690 xg. Total genomic DNA was obtained using

the AxyPrep Bacterial Genomic DNA Miniprep Kit (Axygen Biosciences). The quality of DNA was confirmed using a 1% agarose gel and quantification was performed with the Qubit 2.0 Fluorometer (Life Technologies).

Whole Genome Sequencing, assembling and annotation

Approximately 1 μg of DNA was used to prepare paired-ends libraries with the Nextera XT Sample Prep Kit (Illumina), which were sequenced in the plataform MiSeq (Illumina) using the MiSeq Reagent Kit v3 600 cycle (Illumina) following the manufacturer's instructions. The quality of the raw Illumina sequence reads was initially assessed using FastQC v0.10.1.[50] Seqyclean 1.8.10[51] was used for filtering bases lower than Phred 30 quality. Genome assembly was performed with SPAdes 3.5.0[52] and Platanus 1.2.1 (http://platanus.bio.titech.ac.jp/). The draft genome was annotated with Prokka 1.9.[53] Prediction and annotation of gene clusters encoding enzymes for the production of secondary metabolites were performed with antiSMASH 3.0.4.[54] Draft metabolic models were reconstructed with the RAST annotation system.55 Artemis56 was used for manual annotation and comparisons to previously available protein sequences was carried out using Basic Local Alignment Search Tool (BLAST) from the National Center for Biotechnology Information (NCBI). The identification of motifs and adenylation domains was performed using NRPS Predictor2.[57,58] Information about protein characteristics was obtained by comparisons to the Conserved Domain, UniProt, and Pfam databases.[59–61]

Phylogenetic Analyses

The seven anabaenopeptin gene cluster nucleotide sequences currently available (*Nodularia spumigena* CCY9414, *Nostoc punctiforme* PCC 73102, *Anabaena sp*. 90, *Planktothrix agardhii* NIVA-CYA 126/8, *Planktothrix rubescens* NIVA-CYA 98, *Microcystis aeruginosa* strains PCC 9432 and PCC 9701) were aligned with the anabaenopeptin gene cluster sequence of *S. torques-reginae* ITEP-024 using Clustal W.[62] Evolutionary models were selected by jModelTest 2.1.7[63] and a phylogenetic tree with concatenated genes was reconstructed by the Bayesian inference method with MrBayes 3.2.5[64] using four chains, two separate runs, and 5,000,000 generations. A phylogenetic tree for spumigin genes could not be reconstructed since only one gene cluster (*Nodularia spumigena* CCY9414) is currently available.

Chemical Analysis

Cyanobacterial cells were harvested by centrifugation at 6,350 xg after 18 days of cultivation. Cell pellets were lyophilized, and 10 milligrams of dried material were extracted three times with 2 mL of 70% of methanol solution by probe sonication for 1 minute (Omni Soni Disruptor) and centrifuged at 7840 xg for 10 minutes (Eppendorf, 5804R Centrifuge). The supernatants were pooled together, evaporated to dryness under nitrogen stream (Techal, TE-019 Concentrator), re-dissolved in 1 mL of MeOH:H2O (50%) and subjected to LC-MS/MS analyses. Anabaenopeptin analysis in cyanobacterial extract was carried out in a Shimadzu Prominence LC system (Shimadzu) coupled to a Quadrupole time of flight instrument (MicroTOF-QII; Bruker Daltonics) and equipped with an electrospray ionization interface. Separation was achieved in a RP column (Phenomenex, Luna C18 (2) column, 5 microm, 250 mm x 3.0 mm) at 200 µL min−1 using a linear gradient elution (5–90% B in 50 min). The mobile phases used were (A) water containing 5 mM ammonium formate and 0.1% formic acid, and (B) acetonitrile. The ionization source conditions were as follows: positive ionization, capillary potential of 3,500 V, temperature and flow of drying gas (nitrogen) of 5 mL min-1 and 300 ºC, respectively, nebulizer pressure of 35 psi. The Q/TOF instrument was operated in scan and AutoMS/MS mode, performing MS/MS experiments on the three most intense ions from each MS survey scan. UPLC-QTOF analyses were performed with Acquity I-Class UPLC-Synapt G2-Si HDMS (Waters Corp.) system. 0.2 µL filtered cyanobacterial methanol extract was injected to Cortecst UPLC. C18+ column (2.1 x 50 mm, 1.6 µm, Waters) which was eluted at 40 ÅãC with a flow rate of 0.3 ml min-1 from 5% acetonitrile/isopropanol (1:1, + 0.1% HCOOH) (solvent B) in 0.1% HCOOH to 100% of B in 5 min, kept there 2 min, then back to 5% of B in 0.5 min and finally kept there 2.5 min before next run. QTOF was calibrated with sodium formate giving a calibrated mass range from *m/z* 91.039 to 1178.624. Leucine Enkephalin was used at 10 s interval as a lock mass reference compound. Mass spectral data was accumulated in positive electrospray ionization Resolution Mode at scan range of *m/z* from 50 to 1200. The accurate mass data were processed using Data Analysis 4.0 software (Bruker Daltonics) which provided a ranking of possible elemental formulae (EF) by using the SmartFormulaEditor™.

ASSOCIATED CONTENT

*Supporting Information

Detailed data from bioinformatic, LC-MS, QTOF, and UPLC-ESI-HRQTOF analyses as indicated in the text. This material is available free of charge via the Internet at

http://pubs.acs.org. The genome nucleotide sequence region of 61,333 bp from the *S. torques-reginae* ITEP-024 was deposited in GenBank under accession number KX788858.

References

(1) Welker, M., and von D.hren, H. (2006) Cyanobacterial peptides – nature's own combinatorial biosynthesis. FEMS Microbiol. Rev. 30, 530–563.

(2) Sivonen, K., and B.rner, T. (2008) Bioactive compounds produced by cyanobacteria. In The cyanobacteria: molecular biology, genomics, and evolution. (Herrero, A., Flores, E., Eds), pp. 159–197, Caister Academic Press, Norfolk.

(3) Dittmann, E., Gugger, M., Sivonen, K., and Fewer, D. P. (2015) Natural product biosynthetic diversity and comparative genomics of the cyanobacteria. Trends Microbiol. 23, 642–652.

(4) Murakami, M., Ishida, K., Okino, T., Okita, Y.; Matsuda, H., and Yamaguchi, K. (1995) Aeruginosins 98-A and B, trypsin inhibitors from the blue-green alga Microcystis aeruginosa (NIES-98). Tetrahedron Lett. 36, 2785–2788.

(5) Williams, P. G., Yoshida, W. Y., Quon, M. K., Moore, R. E., and Paul, V. J. (2003) The structure of palau'amide, a potent cytotoxin from a species of the marine cyanobacterium Lyngbya. J. Nat. Prod. 66, 1545–1549.

(6) Tan, L. T. (2007) Bioactive natural products from marine cyanobacteria for drug discovery. Phytochemistry 68, 954–979.

(7) Tidgewell, K.; Clark, B. R., and Gerwick, W. H. (2010) The natural products chemistry of cyanobacteria. In Comprehensive Natural Products II Chemistry and Biologym (Mander, L., Lui, H.-W., Eds), pp. 141-188, Elsevier, Oxford.

(8) Trivella, D. B. B., Pereira, A. R., Stein, M. L., Kasai, Y., Byrum, T., Valeriote, F. A., Tantillo, D. J., Groll, M., Gerwick, W. H., and Moore, B. S. (2014) Enzyme inhibition by hydroamination: Design and mechanism of a hybrid carmaphycin-syringolin enone proteasome inhibitor. Chem. Biol. 21, 782–791.

(9) Kalaitzis, J. A., Lauro, F. M., and Neilan, B. A. (2009) Mining cyanobacterial genomes for genes encoding complex biosynthetic pathways. Nat. Prod. Rep. 26, 1447–1465.

(10) Leikoski, N., Liu, L., Jokela, J., Wahlsten, M., Gugger, M., Calteau, A., Permi, P., Kerfeld, C. A., Sivonen, K., and Fewer, D. P. (2013) Genome mining expands the chemical diversity of the cyanobactin family to include highly modified linear peptides. Chem. Biol. 20, 1033–1043.

(11) Calteau, A., Fewer, D. P., Amel Latifi, A., Coursin, T., Laurent, T., Jokela, J., Kerfeld, C. A., Sivonen, K., Piel, J., and Gugger, M. (2014) Phylum-wide comparative genomics unravel the diversity of secondary metabolism in cyanobacteria. BMC Genomics 15, 977.

(12) Micallef, M. L., D'Agostino, P. M., Al-Sinawi, B., Neilan, B. A., and Moffitt, M. C. (2015) Exploring cyanobacterial genomes for natural product biosynthesis pathways. Mar. Genomics 21, 1–12.

(13) Zhang, Z., Mao, Y., Li, D., Zhang, Y., Li, W., Jia, H., Zheng, J., Li, L., Lu, Y. (2016) Highlevel expression and characterization of two serine protease inhibitors from Trichinella spiralis. Vet. Parasitol. 219, 34–39.

(14) Anas, A. R. J., and Harada, K. (2016) Evaluation of serine protease inhibitors as potent FVIIa-sTF inhibitors in the blood coagulation cascade. Lett. Drug Des. Discov.13, 3–23.

(15) Ersmark, K., Del Valle, J. R., and Hanessian, S. (2008) Chemistry and biology of the aeruginosin family of serine protease inhibitors. Angew. Chem. Int. Ed. Engl. 47, 1202–1223.

(16) Liu, L. W., Jokela, J., Wahlsten, M., Nowruzi, B., Permi, P., Zhang, Y. Z., Xhaard, H., Fewer, D. P., and Sivonen, K. (2014) Nostosins, trypsin inhibitors isolated from the terrestrial cyanobacterium Nostoc sp. strain FSN. J. Nat. Prod. 77, 1784–1790.

(17) Silva-Stenico, M. E.; Rigonato, J., Leal, M. G. Vaz, M. G., Andreote, A. P., and Fiore, M. F. (2012) Non-ribosomal halogenated protease inhibitors from cyanobacterial isolates as attractive drug targets. Curr. Med. Chem. 19, 5205-5213.

(18) Turk, B. (2006) Targeting proteases: successes, failures and future prospects. Nat. Rev. 5, 785–799.

(19) Vijayakumar, S., and Menakha, M. (2015) Pharmaceutical applications of cyanobacteria – A review. JACME 5, 15–23.

(20) Fewer, D. P., Jokela, J., Rouhiainen, L., Wahlsten, M., Koskenniemi, K., Stal, L. J., and Sivonen, K. (2009) The non-ribosomal assembly and frequent occurrence of the protease inhibitors spumigins in the bloom-forming cyanobacterium Nodularia spumigena. Mol. Microbiol. 73, 924–937.

(21) Rouhiainen, L., Jokela, J., Fewer, D. P., Urmann, M., and Sivonen, K. (2010) Two alternative starter modules for the non-ribosomal biosynthesis of specific anabaenopeptin variants in Anabaena (Cyanobacteria). Chem. Biol. 17, 265–273.

(22) Fujii, K., Sivonen, K., Adachi, K., Noguchi, K., Sano, H., Hirayama, K., Makoto Suzuki, M., and Harada, K. (1997) Comparative study of toxic and non-toxic cyanobacterial products: Novel peptides from toxic Nodularia spumigena AV1. Tetrahedron Lett. 38, 5525–5528.

(23) Anas, A. R., Kisugi, T., Umezawa, T., Matsuda, F., Campitelli, M. R., Quinn, R. J., and Okino, T. (2012) Thrombin inhibitors from the freshwater cyanobacterium Anabaena compacta. J. Nat. Prod. 75, 1546–1552.

(24) Mazur-Marzec, H., Kaczkowska, M. J., Blaszczyk, A., Akcaalan, R., Spoof, L., and Meriluoto, J. (2013) Diversity of peptides produced by Nodularia spumigena from various geographical regions. Mar. Drugs 11, 1–19.

(25) Sanz, M., D.rr, F. A., and Pinto, E. (2015) First report of spumigin production by the toxic Sphaerospermopsis torques-reginae cyanobacterium. Toxicon 108, 15–18.

(26) Harada, K.-I., Fujii, K., Shimada, T., Suzuki, M., Sano, H., and Adach, K. (1995) Two cyclic peptides, anabaenopeptins, a third group of bioactive compounds from the cyanobacterium Anabaena flos-aquae NRC 525-17. Tetrahedron Lett. 36, 1511–1514.

(27) Spoof, L., Błaszczyk, A., Meriluoto, J., Cegłowska, M., and Mazur-Marzec, H. (2016) Structures and activity of new anabaenopeptins produced by Baltic Sea cyanobacteria. Mar. Drugs 14, 8.

(28) Horigome, Y., Satake, M., Oshima, Y., Yasumoto, T., Lee, J.-S. (1999) Structure and synthesis of bitter taste peptide from Korean oysters. Tennen Yuki Kagobutsu Toronkai Koen Yoshishu 41st, 409-414.

(29) Tooming-Klunderud, A., Sogge, H., Rounge, T. B., Nederbragt, A. J., Lagesen, K., Gl.ckner, G., Hayes, P. K., Rohrlack, T., and Jakobsen, K. S. (2013). From green to red: horizontal gene transfer of the phycoerythrin gene cluster between Planktothrix strains. Appl. Environ. Microbiol. 79, 6803-6812.

(30) Adiv, S., and Carmeli, S. (2013) Protease inhibitors from Microcystis aeruginosa bloom material collected from the Dalton Reservoir, Israel. J. Nat. Prod. 76, 2307-2315.

(31) Teta, R., Della Sala, G., Glukhov, E., Gerwick, L., Gerwick, W. H., Mangoni, A., and Costantino, V. (2015) Combined LC-MS/MS and molecular networking approach reveals new cyanotoxins from the 2014 cyanobacterial bloom in Green Lake, Seattle. Environ. Sci. Technol. 49, 14301-14310.

(32) Harms, H., Kurita, K. L., Pan, L., Wahome, P. G., He, H., Kinghorn, A. D., Carter, G. T., and Linington, R. G. (2016) Discovery of anabaenopeptin 679 from freshwater algal bloom material: Insights into the structure–activity relationship of anabaenopeptin protease inhibitors. Bioorg. Med. Chem. Lett. 26, 4960–4965.

(33) Schreuder, H., Liesum, A., L.nze, P., Stump, H., Hoffmann, H., Schiell, M., Kurz, M., Toti, L., Bauer, A., Kallus, C., Klemke-Jahn, C., Czech, J., Kramer, D., Enke, H., Niedermeyer, T. H., Morrison, V., Kumar, V., and Br.nstrup, M. (2016). Isolation, co-crystallization and

structure-based characterization of anabaenopeptins as highly potent inhibitors of activated thrombin activatable fibrinolysis inhibitor (TAFIa). Sci. Rep. 6, 32958.

(34) Christiansen, G., Philmus, B., Hemscheidt, T., and Kurmayer, R. (2011) Genetic variation of adenylation domains of the anabaenopeptin synthesis operon and evolution of substrate promiscuity. J. Bacteriol. 193, 3822–3831.

(35) Kehr, J. C., Gatte Picchi, D., and Dittmann, E. (2011) Natural product biosyntheses in cyanobacteria: a treasure trove of unique enzymes. Beilstein J Org Chem. 7, 1622–1635.

(36) Itou, Y., Suzuki, S., Ishida, K., and Murakami, M. (1999) Anabaenopeptins G and H, potent carboxypeptidase A inhibitors from the cyanobacterium Oscillatoria agardhii (NIES 595). Bioorg. Med. Chem. Lett. 9, 1243–1246.

(37) Murakami, M., Suzuki, S., Itou, Y., Kodani, S., and Ishida, K. (2000) New anabaenopeptins, potent carboxypeptidase-A inhibitors from the cyanobacterium Aphanizomenon flos-aquae. J. Nat. Prod. 63, 1280–1282.

(38) Reshef V., and Carmeli, S. (2002) Schizopeptin 791, a new anabeanopeptin-like cyclic peptide from the cyanobacterium Schizothrix sp. J. Nat. Prod. 65, 1187–1189.

(39) Rounge, T.B., Rohrlack, T., Nederbragt, A. J., Kristensen, T., and Jakobsen, K. S. (2009) A genome-wide analysis of nonribosomal peptide synthetase gene clusters and their peptides in a Planktothrix rubescens strain. BMC Genomics 10, 396.

(40) Sano, T., and Kaya, K. (1995) Oscillamide Y, a chymotrypsin inhibitor from toxic Oscillatoria agardhii. Tetrahedron Lett. 36, 5933–5936.

(41) Williams, D. E., Craig, M., Holmes, C. F. B., and Andersen, R. J. (1996) Ferintoic acids A and B, new cyclic hexapeptides from the freshwater cyanobacterium Microcystis aeruginosa. J. Nat. Prod. 59, 570–575.

(42) Zi, J., Lantvit, D. D., Swanson, S. M., and Orjala, J. (2012) Lyngbyaureidamides A and B, two anabaenopeptins from the cultured freshwater cyanobacterium Lyngbya sp. (SAG 36.91). Phytochemistry 74, 173–177.

(43) Koketsu, K., Mitsuhashi, S., and Tabata, K. (2013) Identification of homophenylalanine biosynthetic genes from the cyanobacterium Nostoc punctiforme PCC73102 and application to its microbial production by Escherichia coli. Appl. Environ. Microbiol. 79, 2201–2208.

(44) Kom.rek, K., Kaštovský, J., Mareš, J., and Johansen, J. R. (2014) Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. Preslia 86, 295–335.

(45) Werner, V. R., Laughinghouse IV, H. D., Fiore, M. F., Sant'Anna, C. L., Hoff, C., Santos, K. R. S., Neuhaus, E. B., Molica, R. J. R., Honda, R. Y., and Echenique, R.O. (2012)

Morphological and molecular studies of Sphaerospermopsis torques-reginae (Cyanobacteria, Nostocales) from South American water blooms. Phycologia 51, 228–238.

(46) Dörr, F. A., Rodriguez, V., Molica, R., Henriksen, P., Krock, B., and Pinto E. (2010) Methods for detection of anatoxin-a(S) by liquid chromatography coupled to electrospray ionization-tendem mass spectrometry. Toxicon 55, 92–99.

(47) Vo., B., Bolhuis, H., Fewer, D. P., Kopf, M., M.ke, F., Haas, F., El-Shehawy, R., Hayes, P., Bergman, B., Sivonen, K., Dittmann, E., Scanlan, D. J., Hagemann, M., Stal, L. J., and Hess, W. R. (2013) Insights into the physiology and ecology of the brackish-water-adapted cyanobacterium *Nodularia spumigena* CCY9414 based on a genome-transcriptome analysis. PLoS One 8(3), e60224.

(48) Molica, R. J. R., Oliveira, E. J. A., Carvalho, P. V. V. C., Costa, A. N. S. F., Cunha, M. C. C., Melo, G. L., and Azevedo, S. M. F. O. (2005) Occurrence of saxitoxins and an anatoxin a(s)-like anticholinesterase in a Brazilian drinking water supply. Harmful Algae 4, 743–753.

(49) Gorham, P. R., Mclachlan, J. R., Hammer, V. T., and Kim, W. K. (1964) Isolation and culture of toxic strains of Anabaena flos-aquae (Lyngb.) Breb. Verh. Internat. Verein. Theor. Angew. Limnol. 15, 796–804.

(50) Andrews, S. (2010) FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

(51) Zhbannikov, I. Y., Hunter, S. S., and Settles, M. L. (2015) SeqyClean: a Software Tool for Comprehensive Preprocessing of Sequence Data. Available online at: https://github.com/ibest/seqyclean

(52) Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin A. V., Sirotkin, A. V., Vyyahhi, N., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. (2014) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477.

(53) Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069.

(54) Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., Lee, S. Y., Fischbach, M. A., Müller, R., Wohlleben, W., Breitling, R., Takano, E., and Medema, M. H. (2015) antiSMASH 3.0 – a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res. 43, W237–W243.

(55) Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F., and Stevens, R.

(2014) The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res. 42, D206–D214.

(56) Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A., and Barrell, B. (2000) Artemis: sequence visualization and annotation. Bioinformatics 16(10), 944-945.

(57) Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W., and Hudson, D. H. (2005) Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). Nucleic Acids Res. 33, 5799–5808.

(58) R.ttig, M., Medema, M. H., Bin, K., Weber, T., Rausch, C., and Kohbacher, O. (2011) NRPSpredictor2 — a web server for predicting NRPS adenylation domain specificity. Nucleic Acids Res. 39, 362–367.

(59) Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., Lanczycki, C. J., Lu, F., Marchler, G. H., Song, J.S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., and Bryant, S. H. (2015) CDD: NCBI's conserved domain database. Nucleic Acids Res. 43, 1–5.

(60) The UniProt Consortium (2015) UniProt: a hub for protein information. Nucleic Acids Res. 43, D204–D212.

(61) Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., and Bateman, A. (2016) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44, D279–D285.

(62) Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positionspecific gap penalties and weigh matrix choice. Nucleic Acids Res. 22, 4673–4680.

(63) Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9, 772.

(64) Ronquist, F., and Huelsenbeck, J. P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Table 1. Conservation of the eight adenylation domain binding pockets of Apt of *S. torques-reginae* ITEP-024 and other cyanobacteria

Strain/Binding pocket	**AptA1*-A_1**	**AptA1*-A_2**	**AptA**-A_1**	**AptA**-A_2**	**AptB-A**	**AptC-A_1**	**AptC-A_2**	**AptD-A**
S. torques-reginae ITEP-024	-	-	DAFFLGVTYK Ile 90 %	DTEDIGSVVK Lys 70 %	DGLWLGGVFK Val 80 %	DLGFTGCVTK Leu 60 %	DLFNNALTYK Ala 100%	DAWTIGGICK Phe 80 %
Anabaena sp. 90	DVEDIGSVAK Lys 70%	DTEDIGSVIK Lys 70%	--STIAAVC- Tyr 100 %	---------- Lys 70 %	-A-------- Val 90 %	---------- Leu 60 %	---------- Ala 100%	-----A-V-- Phe 100 %
N. spumigena CCY 9414	-	-	---------- Ile 90 %	------T--- Lys 70 %	-MWFM---I- Val 70 %	---AI---I- Asn 60 %	--GFSGCVT- Leu 60%	-V---TA--- Phe 80 %
P. agardhii NIVA-CYA 126/8	-	-	-VESI-AIA- Trp 70 %	-A-------- Lys 80 %	-A--I----- Val 90 %	---------- Leu 60 %	--GFTGCVT- Leu 60%	--FFL-VTF- Ile 100 %
P. rubescens NIVA-CYA 98	-	-	-VESI--IA- Orn 60 %	-A-------- Lys 80 %	-A--I----- Val 90 %	---------- Leu 60 %	---------- Ala 100%	-----A-V-- Phe 100 %
N. punctiforme PCC 3102	-	-	--WTIAAIC- Phe 100 %	-A------I- Lys 80 %	-A-FM-V--- Phe 70 %	---TI---I- Asn 60 %	-ILQLGVIW- Gly 100%	-LGFT-CVN- Leu 60 %
M. aeruginosa PCC 9432	-	-	-VHDI-AIE- Arg 70%	-A-------- Lys 80%	-A--I----- Val 90%	----A----- Lys 60%	---------- Ala 100%	-----A-V-- Phe 100%
M. aeruginosa PCC 9701	-	-	-VESI---A- Orn 60%	-A-------- Lys 80%	-A--I----- Val 90%	----A----- Lys 60%	---------- Ala 100%	-----A-V-- Phe 100%

The adenylation domain is responsible for recognition and activation of amino acids in the anabaenopeptin. The probability of incorporation for each amino acid is showed.
**AptA1-A_1 and AptA1-A_2 are found only in Anabaena sp 90. **AptA-A_1 and AptA-A_2 are AptA2 in Anabaena sp 90.*

Table 2. Conservation of the four adenylation domain binding pockets of Spu of *S. torques-reginae* ITEP-024 and *N. spumigena* CCY 9414

Strain/Binding pocket	SpuA-A	SpuB-A_1	SpuB-A_2	SpuB-A_3
S. torques-reginae ITEP-024	GIFWHGGSG- α-hydroxy-isocaproic acid 60 %	DASTIAAVCK Tyr 100 %	DVQFIAHAVK Pro 90 %	DVETTGAVTK Arg 70 %
N. spumigena CCY 9414	ALL-IAA--- α-hydroxy-isocaproic acid 50 %	-LAFTGC-T- Leu 60%	---------- Pro 90%	---------- Arg 70%

The adenylation domain is responsible for recognition and activation of amino acids in the spumigin. The probability of incorporation for each amino acid is showed.

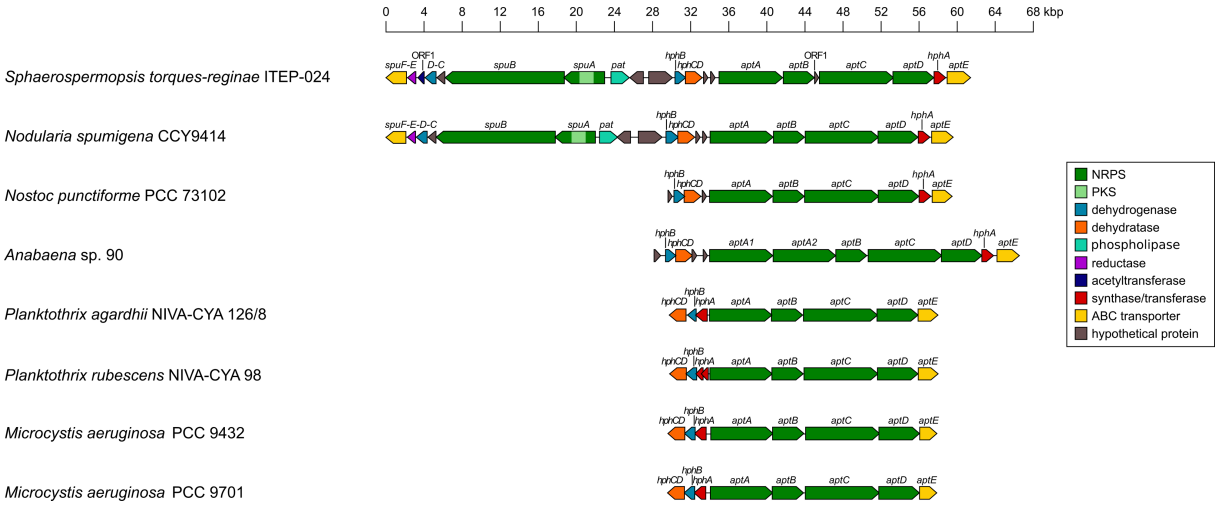


Fig. 1. Organization of the spumigin (*spu*) and anabaenopeptin (*apt*) biosynthetic gene clusters from strains of different cyanobacterial genera. The previously nostocalean *aptF* gene, which encodes an ATP binding cassette (ABC) transporter, was renamed *aptE*.

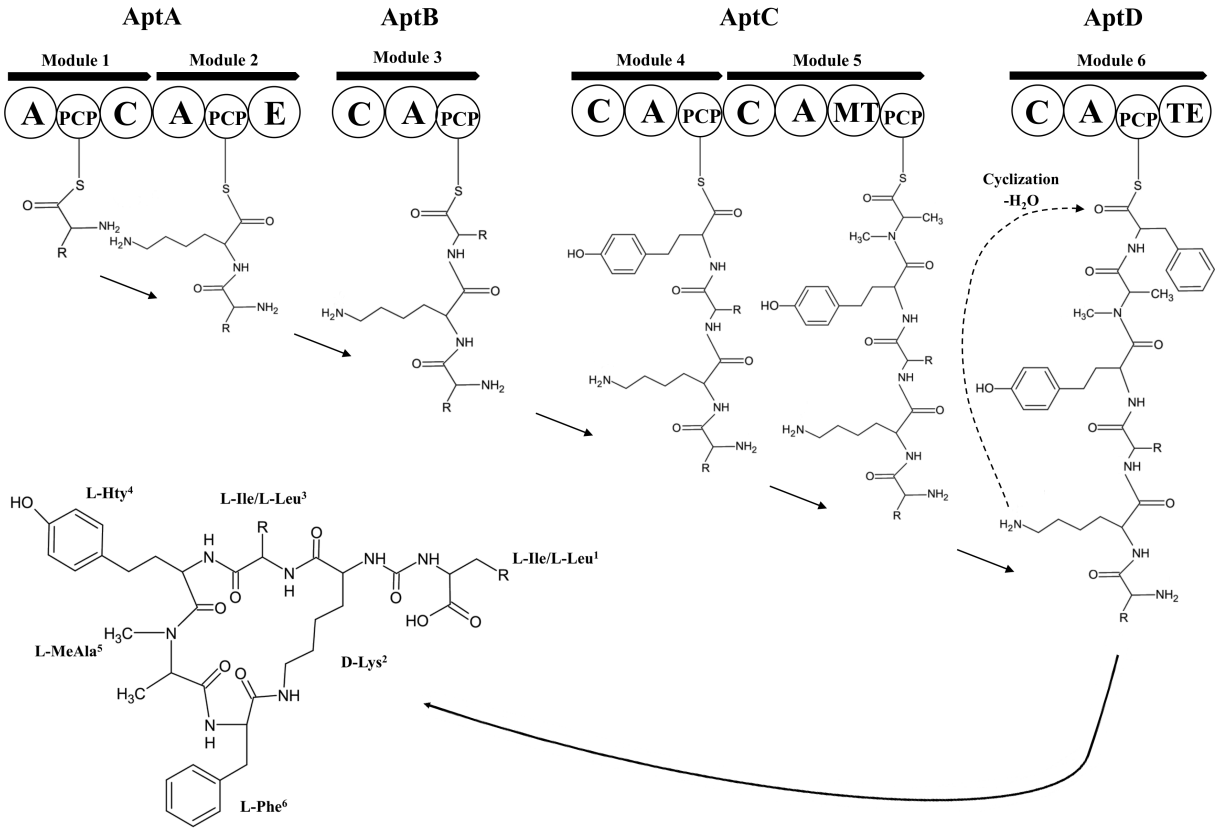


Fig. 2. Map of apt gene cluster and proposed pathway of anabaenopeptins and its chemical structure produced by *S. torques-reginae* ITEP-024. The substrate amino acid L-lysine is isomerized to D-lysine by the epimerase domain in AptA. The step of how the ureido bond is formed is still unclear. Abbreviations: A, adenilation; PCP, peptidyl carrier protein; KR, keto reductase domain; C, condensation; E, epimerase domain; MT, methyltransferase domain; R, reductase domain; TE, thioesterase domain; TD, thioester reductase; L-Ile, L-isoleucine; L-Leu, L-leucine; D-Lys, D-lysine; L-Hty, L-homotyrosine; L-MeAla, L-methylalanine; L-Phe, Lphenylalanine.

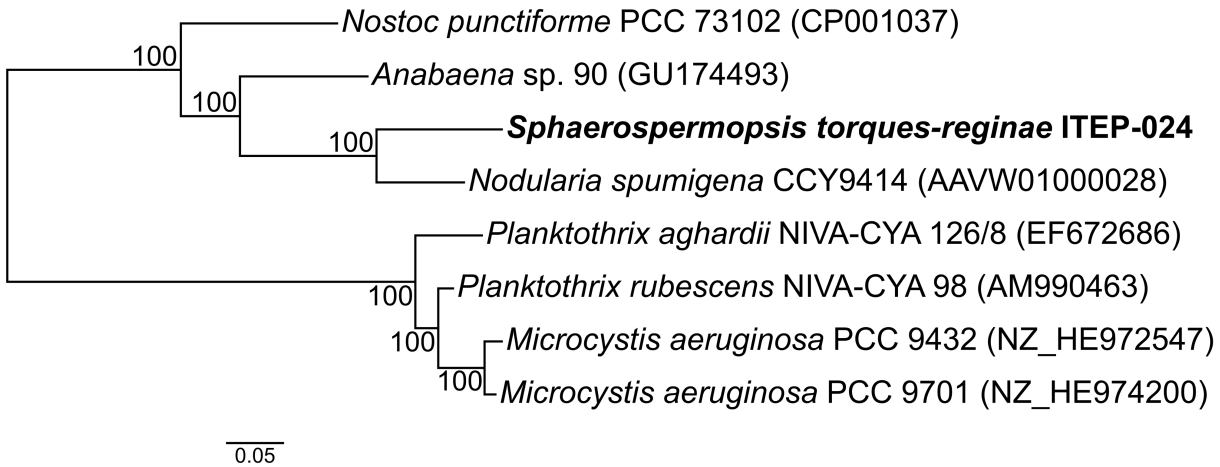


Fig. 3. Phylogenetic tree of concatenated nucleotide sequences of the eight complete gene clusters encoding anabaenopeptins using the Bayesian inference method. The strain studied (ITEP-024) is highlighted in bold. Posterior probabilities are represented on each node.

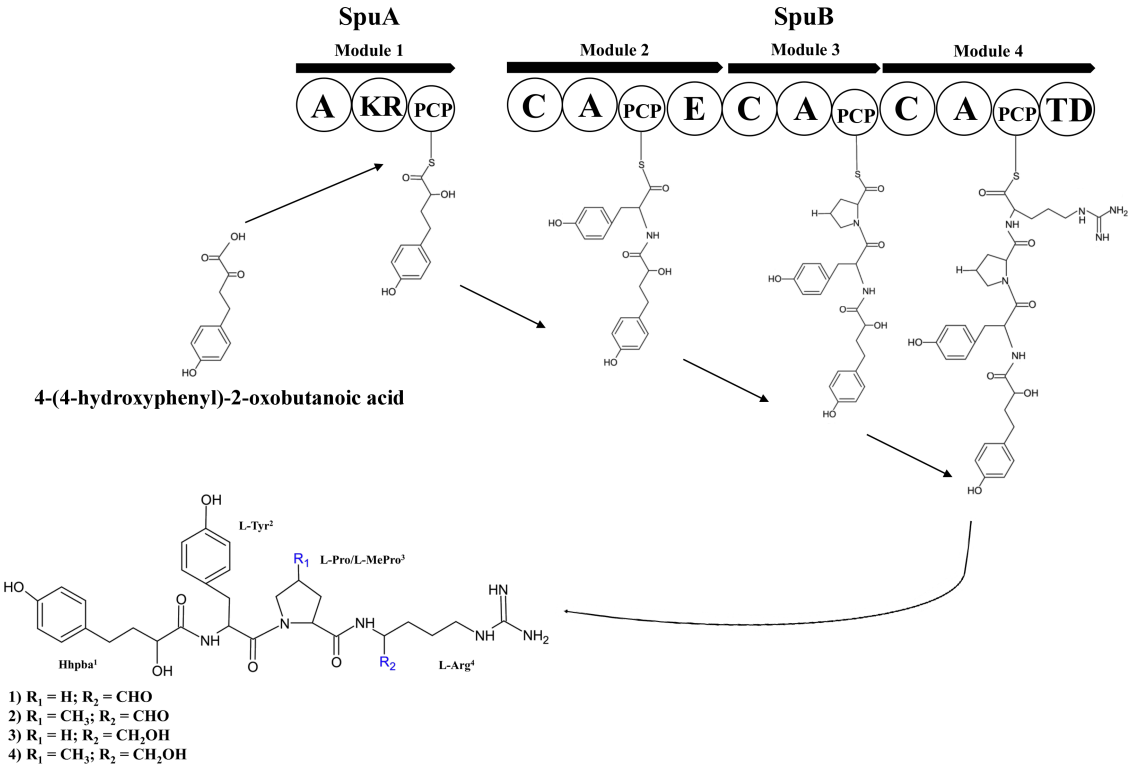


Fig. 4. Map of spu gene cluster and proposed pathway of linear tetrapeptide spumigins and the chemical structure of the four isoforms produced by S. torques-reginae ITEP-024. The variants differ in the presence of Pro or MePro in position three and in the presence of Argol or Argal. Abbreviations: A, adenilation; PCP, peptidyl carrier protein; KR, keto reductase domain; C, condensation; E, epimerase domain; MT, methyltransferase domain; R, reductase domain; TE, thioesterase domain; TD, thioester reductase; Hhpba, hydroxy-(hydroxyphenyl) butyric acid; L- Tyr, L-tyrosine; L-Pro, L-proline; MePro, (2S, 4S)–4–methylproline; Argol, argininol; Argal, argininal.

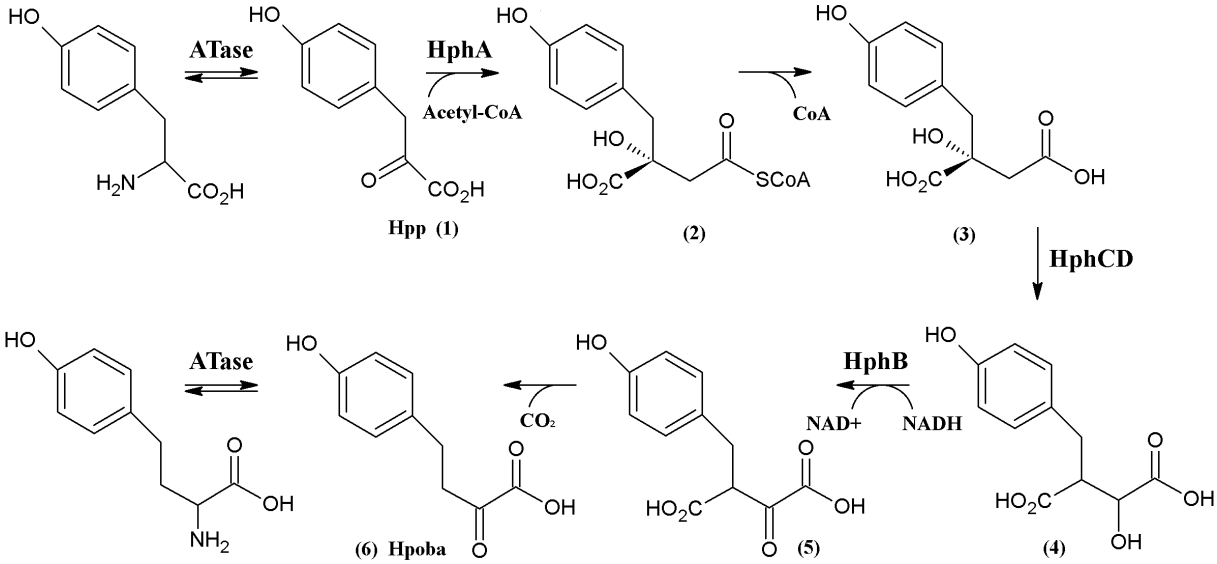


Fig. 5. The proposed 4-(4-hydroxyphenyl)-2-oxobutanoic acid (Hpoba) biosynthetic pathway mediated by HphABCD enzymes and aromatic aminotransferase (ATase). L-Tyr is converted into 4-hydroxyphenylpyruvic acid via a transamination reaction mediated by an aminotransferase. 4-hydroxyphenylpyruvic acid is condensed with acetyl-coenzyme A (CoA) catalyzed by HphA, and the resulting thioester will be hydrolyzed by a spontaneous reaction leading to the intermediate (3). (3) is converted into (4) via isomerization of a hydroxyl group mediated by HphCD enzyme. The hydroxyl moiety of (4) is oxidized by HphB enzyme leading to the intermediate (5). The intermediate (5) followed by decarboxylation spontaneous reaction is converted into Hpoba. Finally, Hpoba can be converted into L-Hty via a transamination reaction mediated by an aminotransferase (ATase).